



Deontic Intentionality

Citation

Gabor, Zachary. 2024. Deontic Intentionality. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37378984>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

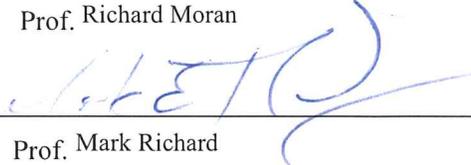
The undersigned, appointed by the
Department of Philosophy
have examined a dissertation entitled
Deontic Intentionality

presented by Zachary Gabor

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature 

Typed name: Prof. Richard Moran

Signature 

Typed name: Prof. Mark Richard

Signature 

Typed name: Prof. Susanna Siegel

Signature _____

Typed name: Prof.

Signature _____

Typed name: Prof.

Date: 5/6/24

Deontic Intentionality

A dissertation presented

by

Zachary Gabor

to

The Department of Philosophy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Philosophy

Harvard University

Cambridge, Massachusetts

May 2024

© Zachary Gabor 2024

All rights reserved

Deontic Intentionality

Abstract

I aim to vindicate, in naturalistic terms, the idea that our beliefs ‘aim at truth’ in a strikingly strong sense: we are subject to the *Truth Norm* (“TN”) on our beliefs, whose content is *one ought to believe p iff p*. The first chapter motivates this aim. Deploying a twist on Lewis Carroll’s parable of Achilles and the Tortoise, I argue that our practice of forming shared consensus about what is the case through conversation embodies, in its very structure, an understanding of ourselves as subject to TN. We cannot seriously contemplate a way of life in which this practice is not available, so this leaves us two choices: either TN really does govern our beliefs, or we are driven to the skeptical conclusion that we are stuck fundamentally misunderstanding ourselves. Avoiding this skeptical conclusion is at least worth a try. I further argue that a naturalistic account of TN would constitute progress toward a naturalistic understanding of the relation of the objective to the subjective. Finally, I indicate a general strategy for overcoming the obvious objections to TN: objections from the existence of truths for which we have no evidence, truths which would be cumbersome or distressing to believe, and so on. These objections indeed point to problems with the most immediately available ways of believing the truth. But it may be that ψ -ing is the most immediately available way of ϕ -ing, and that you ought not to ψ , but that you nonetheless ought to ϕ .

The remainder of the dissertation aims to give a naturalistic account of the relevant sorts of beliefs and requirements according to which just such a situation obtains with respect the Truth

Norm and our beliefs. I propose notions of requirement and belief built of from teleological building blocks and to argue that they fit together to vindicate TN in the human form of life, and to give a naturalistic account of the relevant teleological building blocks.

First, the building blocks: the project requires a notion of purpose that, while naturalistic, is liberal enough to vindicate as reality the appearances of purposiveness in both biological and social organization. I develop a notion of purpose answering to this requirement. I argue that, whereas critics have held that such efforts rely on an oversimplification of empirical reality, it is actually these critics who commit a converse kind of oversimplification: they oversimplify, and thus unduly constrain, what it would take for the empirical facts to ground the reality of the widespread appearance of teleology in the biological and social world.

Next, the notions of belief and requirement. A belief is a state of preparedness (a “disposition, teleologically enriched”) to produce indicative signals with a given content. To have the function of an indicative signal with content *p* is, in turn, to be meant to adapt those elements of the recipient’s teleological organization which are sensitive to whether-*p* to the contingency that-*p*. To be subject to a hypothetical imperative *if you’re A-ing, you ought to B* is to be in a situation in which there is a substantive opportunity to B in keeping with the normal functioning of your capacity to do something that plays the role of making and recognizing an ought-judgment. The job-description of such an ought-judgment comprises functioning as an indicative signal with content *A-ing calls for B-ing* – a notion which may be glossed as the circumstance that best efforts to A entail efforts to A by B-ing – and an imperative signal prompting or ensuring B-ing. TN will be functionally like a categorical imperative, a hypothetical imperative with the antecedent *if you’re leading a human life...*

Last comes the argument for the Truth Norm. It involves working with precise formulations of the notions merely glossed above, but the intuition is this: there are limits on the extent to which

nonhuman animals are subject to something akin to TN as a norm on their representational states, deriving from the limitations on the range of ends around which their lives are teleologically organized. There is only so much a non-human animal might be able to do to get true beliefs, or to put them to good use. *We*, on the other hand, make our living by forming and working in social arrangements of potentially endless variety. Our way of life involves availing ourselves of the efforts of the groups we are part of to get true beliefs, and it also involves expressing our beliefs to further the group efforts we are part of by adapting them to the facts. Thus, I argue, there are not substantive boundaries on the truths we might hope to attain or the uses to which we might put them, and so, we are subject to TN.

Table of Contents

Title Page – i

Copyright – ii

Abstract – iii

Table of Contents – vi

Dedication – vii

Introduction – 1

Chapter 1: The Manifest Normative Character of Belief – 5

Chapter 2: Teleology on the Heath – 34

Chapter 3: A Guided Tour of the Teleological Landscape – 67

Chapter 4: Universality of Repertoire and Truth-Normativity – 134

Bibliography – 164

For Debbie and Ricky, my mom and dad.

Introduction

It is often said that intentional notions – those that involve representation or aboutness – seem to form a circle. Perhaps you could say what it is for an utterance to mean something if you know what a belief or an intention is. And perhaps you can say what it is to follow a rule if you know what it is to think or utter something meaningful. But it is hard to define any one without appeal to some others. An attempt to do so is an attempt to “break into the intentional circle.”

One way to interpret the rhetoric of “breaking in” is as a metaphorical allusion to a heist attempt. And that interpretation occasions the question: *just what will we make off with if we succeed?* This question turns out to be difficult. Any attempt to define an intentional notion – say, that of following a rule – in non-intentional terms may face the criticism that it simply meanders around the outside of the circle, providing some condition that generally covary with the presence of rule-following but do not capture its essence. One would like a test – some criterion by which we would be able to tell whether a putative definition of an intentional notion in non-intentional terms is the real article.

I adopt an idea from Sellars, that among the most valuable jewels locked within the intentional circle, and one of the hardest to make off with, is the normative value of truth. Sellars’s idea about the form this quarry took was that truth amounted to semantic assertability (1968) – that what it was for an utterance to truly represent the world was for the rules governing what it is proper say to permit asserting it. One difficulty with this idea is that it requires, if it is to be cashed out, a way of distinguishing between the *semantic* rules determining what one ought to say from other rules about what one ought to say, such as those having to do with pragmatics or etiquette. If the goal is to say what it is for an utterance to be true without appeal to another intentional notion, then to say

that the semantic rules are just the rules that pertain directly to the meaning of what is said, is just to take one step around the circle, just what we are trying to avoid.

In this dissertation, I flip two things about Sellars's approach to the relationship between truth, normativity, and representation. Instead of trying to do things in terms of *permissions*, that pertain to the *actualities* of assertion, I will try to do things in terms of *requirements* on the *potentialities* which under the right circumstances manifest in honest assertions, i.e. beliefs. The normative element of intentionality, on the view I advocate here, centers around the requirement that one believe what is true. This approach has the advantage that it faces no need to separate out a specifically semantic stratum of normativity.

It faces the disadvantage that the idea that one ought to believe each of and only the true propositions seems wildly implausible at first pass. What about useless true beliefs, distracting true beliefs, true beliefs that would take too much effort to obtain, or true beliefs in the face of massive misleading evidence? As implausible as it may seem, I will argue that the idea that belief is normatively related to truth in this rigid way is an essential part of the conception of our subjectivity that we live out. In chapter 1, I try to argue that the very structure of our practices of forming shared agreement presuppose that one ought to believe P if and only if P.

The remainder of the dissertation is an essayistic attempt to make off with this prized jewel. It has the dual goals of constructing a notion of belief that captures core dimensions of our subjectivity, and of arguing that this notion is subject to the Truth Norm *believe P iff P*.

The main ingredient in terms of which I will attempt to construct a notion of belief is the (intensional but not intentional!) notion of explanation denoted by the "commonsense metaphysics" meaning of the English word "because." In chapter 2, propose a notion of teleo-function or

purpose, defined in terms of “because” as well as, other commonsense metaphysics words, such as “conducive.”

In chapter 3, I define notions of belief and requirement, and various other attendant notions that are necessary as intermediates, in terms of the notion of function from the previous chapter, as well as a further smattering of “because” and other commonsense metaphysics words. The notion of belief I propose embodies a kind of “teleologically enriched behaviorism” on which believing P is a matter of being prepared, not merely disposed, to communicate that P in certain circumstances. I advance a notion of something’s being required within the context of an activity that aims to take the idea of a instructing or commanding someone in the conduct of activity, and abstracting away from whatever the idiosyncratic preferences of the instructor or commander may be to focus on what remains after these have been jettisoned.

In chapter 4, I argue for the key claim that I suggest is necessary to establish that the notion of belief I have suggested is subject to Truth Normativity, which I call the Universality Thesis. This is the thesis that we cannot place any substantive limits on the kinds of strategies or abilities with which we are to face the challenges that arise in the course of our functioning. I conclude with some elaboration of and observations about the resulting picture, on which belief is normatively aimed at truth.

Chapter 1: The Manifest Normative Character of Belief

The phenomenon of our subjectivity is distinctive and remarkable. Unlike most other parts of the world, we are parts of the world with points of view on the world. This dissertation focuses on one of the most salient, distinctive, and philosophically puzzling features of these points of view: they involve convictions about how things are. Throughout, I will refer to these convictions as “beliefs,” but because this term is used in more than one way by philosophers, I should be clear what I mean by it. Beliefs in this sense are “full” beliefs, not degrees of confidence. They are conscious in Block’s (2002) “access” sense: they are available for the person who possesses them to think explicitly about. They are the sorts of things that, in academic circles, we refer to as our “views.” To put it in an imprecise, but hopefully evocative way, your beliefs, in my intended sense, are the world, according to you. There is no world according to Mount Vesuvius; there is no world according to Jupiter; and if there is a world according to any specimen of *e. coli*, it is surely in some vastly different sense than that in which there is a world according to you or me. The goal of the dissertation is to go some way toward explaining what exactly it is for us to have beliefs in this sense and how it is possible that we do.

§1 – The Strategy: A Very Big Difference

To begin with, I want to say something about the dissertation’s strategy for doing this and about the role of the first chapter within that strategy.

One common approach to providing an account of X is to start by enumerating a handful of things – maybe three or four – that an account of X should do, then providing an account that does each of the things you’ve enumerated.

There are a few things that should be true of your enumerated criteria if your argument is going to be convincing. First, the criteria probably shouldn’t be too easy for an account of X to meet. Otherwise, it won’t be a distinguishing virtue of your account that it meets these criteria

because it doesn't meet them. Second, the criteria should be *possible* for an account of X to meet jointly. Otherwise, it won't be a distinguishing virtue of your account that it meets these criteria since it won't meet them. Finally, it should be defensible that there is a significant connection between your proposed and an antecedently significant concept of X. Otherwise, opponents of your argument may construe your insistence that an account of X must meet these criteria as a matter of your idiosyncratic conception of X, and then proceed with an alternative account of X that doesn't meet your criteria.

This last *desideratum* could easily be misinterpreted overly strictly, and I want to guard against that explicitly. What needs to be true, for your argument to be any good, is that your criteria are connected with *an* antecedently interesting concept of X, not necessarily the *only* such concept. For if your opponent objects to your account of X that your criteria are connected to the essence of *one* concept of X, and that their account deals with X in another sense, this doesn't necessarily mean that there is any problem with your account or with your argument for it. There is only something wrong with your argument if your criteria come from a conception of X that is *idiosyncratic*; it is no problem if they merely come from the concept of X in what turns out to be one sense of 'X' among others.

I want to pursue something like the routine approach I have just described above. But instead of enumerating a handful of things that an account of belief should do, I am going to enumerate just one criterion that it must meet and then provide an account of belief in keeping with that single criterion. The purpose of this first chapter is to argue for that criterion as a standard for accounts of belief in the relevant sense. This criterion will be a requirement that the account of belief explain what I will call the Manifest Truth-Normativity of belief, which is a thesis about what

might be called the “manifest normative character” of there being a way the world is, according to you, in the way that there is for us.

This single-criterion strategy might appear on its face dispreferable to the version that proceeds from a handful of criteria. Isn't it a much stronger argument for the view if you can point out four good things that your account does instead of just one? But of course, it really all depends on how important the one thing is as compared to the four. It matters how much distinction an account would achieve for meeting the one criterion as opposed to how much it would achieve by meeting the four. And it matters how dispensable or indispensable it is to the concept of X that the account meet the one criterion or that it meet the four criteria, respectively.

My approach to motivating an account of belief is focused on a single criterion that does well on the first test. As I will discuss below, it will take quite a lot for an account of belief to meet the single criterion I suggest in a satisfying way. So if I can give an account of belief that meets it, it will be a distinguishing virtue of that account. My criterion for an account of belief will also do well on the second test, or so I will argue in this chapter. Specifically, I will argue that a conception of belief as characterized by manifest truth-normativity is indispensable, as it is implicit in the very structure of the practice of reasoning. If that is right, then it will be very important indeed to have an account of the notion of belief that can explain Manifest Truth-Normativity. That will be the main claim of the chapter as a whole.

But before proceeding to that argument, I want to say something more general about why questions about the natures of mental properties, my question among them, can tend to be suited to the approach that focuses on one big criterion for an account rather than several smaller ones. The question of how non-thinking stuff can constitute thinking stuff seems like a question about a big difference. Here we have some stuff *viz.* everything in the universe prior to the emergence of the

first thinking thing; and this stuff is all one way *viz.* unthinking. Lo and behold, in a certain arrangement, some of this stuff is a very different way (thinking). An adequate account of the mental properties involved with thinking must get the Very Big Difference right.

I think that one can see the appeal of conceptual questions about mental properties by focusing on a single major criterion in the dialectic of many counterexample-based objections to accounts of mental properties that are in one or another way “reductionist. To take one example among many, consider Ned Block’s “China Brain” thought experiment from “Troubles with Functionalism” (1978) as an example among many others. Block asks the reader to imagine that the population of China were functionally arranged so that each did the job of a single neuron in a replica human brain used to control a robot. Block thinks that there would be no sensing going on in such an arrangement. If that is right, then there must be something wrong with functionalist accounts of those mental properties connected with our sentience. To feel the grip of Block’s proposal, it seems to me, is not to suspect that functionalism checks only three of five boxes, or some such. The putative counterexample is supposed to evoke the intuition that there is a big gulf between things with the mental properties to be explained, and that functionalism cannot show us how to cross it.

Now I don’t aim to convince readers of this chapter that I have identified The Mark of the Mental; I am not trying to capture the whole essence of mentality in one criterion. There are a few reasons for this, but most important among them is that the idea that there is such a thing as *the* one and only mark of the mental is a big assumption and one that I’m not prepared to make. In the first place, it may be that there is not one Very Big Difference between the mental and the non-mental, but two or three or five such differences, none of which is clearly reducible to the others. I’m aiming

to find one VBD and produce a plausible account that meets it. Whether there are other VBDs, and whether my account can meet them isn't something I will address.

Second, it may be that there are different genera of mental properties, and although all of these genera are classified as genera of *mental* properties, each genus has its *own* Very Big Difference associated with it; things exhibiting mental properties of one genus are thereby very different from everything else in one specific way; things exhibiting mental properties of another genus are thereby very different in another specific way. For instance, to have knowledge properties (i.e. to know things), might be one way of being very different from everything that has no knowledge properties, and to have sensational properties (i.e. to have sensations) might be a *different* way of being very different from everything else. In such a case there would be, as it were, the mark of the *knowing* and the mark of the *sensing*; neither of which could be construed as *the* mark of the mental.

So I want to identify and argue for just *one* of perhaps several main very big differences between things exhibiting one *aspect* of mentality and everything else. The kind of mental property in question will be roughly speaking, the intentional aspects of subjectivity. In rough terms, the aspect of thinking things like us that there is such a thing as the world according to it.

§2 – Manifest Truth-Normativity

When something has the property of believing that the cat is on that mat – the property that the world, according to them, is one in which the cat is on the mat – it has a property with that has something particular to do with the cat's being on the mat. What must having this property have to do with the cat's being on the mat for the property to make up part of a subjective point of view?

The idea behind Manifest Truth-Normativity is that one natural place to start is with the felt *directedness* of views about how things are. Views are the products of attempts to figure out how

things are, and they figure in future such attempts. A central feature of beliefs seems to be what we characteristically try to do with them in the course of these attempts: get the true ones. Thus, it seems that one of the main things the property of believing the cat is on the mat has to do with cat's being on the mat is this: the cat's being on the mat seems to be the *success* condition for believing. Belief seems to "aim at truth," as it is sometimes put. If an account of belief cannot get this – the apparent normative character of having a point of view on the world like we do – right, then it is inadequate as a characterization of what gets us across the gulf between everything that doesn't exhibit this kind of subjectivity and us, who do.

The idea that belief seems to aim at truth is rather imprecise. The thesis Manifest Truth-Normativity will be my attempt to make it precise. First, Truth-Normativity is the claim that truth is the norm of belief in the following strong sense:

Truth-Normativity: For any p , one ought to believe p iff p .

Manifest Truth-Normativity is a claim about the "logic" of a notion of belief that is presupposed in our practices.

Manifest Truth-Normativity: The process of reasoning structurally presupposes that our belief-properties in the right sense of "belief" are subject to truth-normativity, i.e. that we ought to have them iff they are true.

The name I've chosen for MTN is intended to evoke Sellars's notion (1962) of "the manifest image of man in the world." The manifest image is not the naïve conception either of ourselves or of the world; rather it is a conception of the world that is the product of refinement *via* rational scrutiny, *but* all the same beholden to the characteristic contours of a human life. For instance, the manifest image is an image of a world populated by persons, because one cannot lead a

characteristically human life without (at least implicitly) thinking of oneself and one's fellows as persons.

Something like this is what I want to claim about the image of ourselves as enjoying beliefs subject to Truth-Normativity. The contention is not that Truth-Normativity is obvious to the person on the street; it is that the person on the street is committed to it by the intellectual framework that enables her to lead the life of a human being, and that this commitment is capable of excavation *via* rational scrutiny. In a phrase, it holds that the principle of Truth-Normativity on belief is a way of saying something that is presupposed in what we do.

Notice that if Manifest Truth Normativity is correct, and especially if it is correct because Truth-Normativity is correct, that gives us a nice, neat, and remarkably simple preliminary characterization of what a belief has to do with what it is a belief of – what its “aboutness” or world-directed character consists in. The worldly condition that a belief is about or directed at is the worldly condition that delineates between (what appear to be) the circumstances in which it should occur and those in which it should not. This is not to say that any property that meets this condition is *ipso facto* a belief property. Over the course of the dissertation, I will elaborate a view on which there is a good deal more to say about what the property of believing P has to do with P than this one normative fact. But all the same, if Manifest Truth-Normativity can be defended, and especially if it can be defended in a vindicatory way that support Truth-Normativity, it provides a neat explanation of the source of the temptation to say that the intentionality of beliefs consists in their being “directed at” or “answerable to” worldly conditions.

The aim of the chapter is to argue that there is a concept of belief such that any adequate account answering to it must make sense of Manifest Truth-Normativity, i.e. of Truth-Normativity being structurally presupposed in reasoning. But it deserves immediate comment that the underlying

putative norms on belief, those given by Truth-Normativity, seem implausibly overdemanding. Truth-Normativity says that if *p* is true, you should believe it. It says so, for instance, even when you have strong evidence against *p*, and even when you can't foresee any use for believing *p*. These apparently prohibitive difficulties for accepting Truth-Normativity will be the focus of much of the discussion of subsequent chapters. But for now, it bears emphasizing that the goal of the *chapter* is not to vindicate Truth-Normativity itself as a claim about this kind of belief. If the argument of this chapter succeeds, then there are still three possibilities with respect to the notion of belief under consideration.

First is the possibility that no adequate account of belief can be given that explains the truth even of *Manifest* Truth-Normativity. That would mean that it turned out that there was nothing at all like the points of view that our reasoning practices presuppose, and presuppose to be governed by truth-normativity, after all. This would be a puzzling and apparently unstable skeptical conclusion, but the argument here isn't meant to rule it out.

Second, it could be that an account of the relevant sort of belief can be given, and that the correct account explains Manifest Truth-Normativity as the *false* appearance of Truth-Normativity governing beliefs. In that case it would turn out that the points of view we presuppose to exist in our reasoning practice did exist, but that they are not subject to Truth-Normativity, as we necessarily presuppose them to. This sort of account would construe us as confused in a fundamental way about the nature of our own minds.

Third is the possibility that an adequate account of beliefs of the relevant kind can be given that explains Manifest Truth-Normativity in a vindicatory way, as a matter of our reasoning *correctly* presupposing that we have views and that these views are subject to Truth-Normativity. In the remaining chapters, I will pursue this third possibility, aiming to advance an account of belief that

explains Manifest Truth-Normativity by vindicating Truth-Normativity for the beliefs in my account. But this will be a tall order; Truth-Normativity seems to place an implausibly demanding normative constraint on believing. In this chapter, I will leave the appearances that Truth-Normativity is just too demanding and can't be true untouched. My goal is just to excavate an additional appearance to the *contrary* – an appearance that Truth-Normativity *must* be true – that I claim is inextricably tied up in the structure of the activity of reasoning.

§3 – What the Hare told Helen: An Argument for Manifest Truth-Normativity

The basic idea behind the argument for Manifest Truth-Normativity is relatively simple. Reasoning appears to be a directed process. Reasoning about whether *p* results in a belief-state, either of believing *p*, believing not-*p*, or not believing either. But it is in the nature of reasoning to only treat considerations as relevant in reasoning about whether *p* if they seem to be relevant to *p*'s truth. So it is in the nature of reasoning to treat believing the truth as the goal, i.e. to treat it as the outcome which ought to occur.

I will spell out this thought in detail in the form of a commentary on an illustration. The illustration is a twist on Lewis Carroll's parable of Achilles and the Tortoise (1895). In 3.1 I summarize Carroll's fable and extract a moral from it. In 3.2, I provide the twist and state the moral that I will argue it supports. 3.3 provides the argument that it does support this moral.

3.1 – Carroll's Fable of Achilles and the Tortoise

The *dramatis personae* of Carroll's fable are Achilles – the voice of reason – and a tortoise. Although the tortoise is in most respects apparently reasonable and competent as an interlocutor, there is one key aspect of the practice of reasoning that he just does not understand,¹ and this

¹ Actually, to be more precise, Carroll's fable has the tortoise simply *pretending* not to understand the relevant point, in order to frustrate Achilles, who has not alighted on the way to explain it to him. As far as I can tell, the fable seems to

prevents him from being able to reason fruitfully with Achilles. Achilles is armed with five things: slate, chalk and three claims, abbreviated A, B, and Z, where the first two entail the third. With these armaments, he attempts to write something down that will show the Tortoise that logic compels him to accept Z, supposing he accepts A, and B. Achilles writes down (C): if A and B are true, then Z must be true. The tortoise accepts C with equanimity, but then asks why accepting A, B, and C require him to accept Z. Thus is Achilles sent off to the races.

A moral I draw from Carroll's story – and I am not supposing that this is the only thing that it shows, or that it is exactly what Carroll intended it to show, but just that it pretty well *does* show it – is the following:

Moral (A&T): In order to convey by *saying* (writing down) all that one needs to know in order to reason logically, one must be able to make imperative-mood statements.

The tortoise appears to accept and understand each conditional Achilles writes down, but none of them supply what the tortoise is (pretending to be) missing. Apparently, Achilles does not think to, or does not know how to, or does not think he is allowed to write down the *imperative*, “if you accept premises of the form X, Y, and if X and Y are true, then Z must be true, and if you are not prepared to revoke any of these acceptances, accept Z!”² If Achilles were to write this down, the tortoise would not be able to go on playing the same game.

get the point across either way, and Carroll has the tortoise merely pretending not to understand for literary and rhetorical reasons, not philosophical ones. I mostly elide this point in my discussion of Carroll's fable and elide it entirely in my twist on the fable because I take it not to make a difference for my purposes and doing so allows me to make my points in a more straightforward way.

² “Does not think he is allowed to” seems to me to be the best reading of the dialogue as Carroll explicitly constructs it. What the tortoise says seems to heavily implicate that Achilles is only to write down something that can serve as a premise, i.e. an assertion, not an imperative. But again, I don't see that this is philosophically important.

To be sure, the tortoise might say that he accepts this command about what he must do and still not draw the conclusion. But this would be a very different kind of intransigence. All that *this* intransigence would go to show is:

an interlocutor can tell you they accept what you say without having actually done so, either because they are lying to you, or because they don't understand you.

As long as Achilles only employs conditionals connecting the premises and the conclusion, the tortoise can appear to accept and understand everything Achilles writes down without accepting the conclusion. But once Achilles has written down the command, if the tortoise holds his ground, he will appear to be merely mouthing an acceptance of the imperative; depending on the situation Achilles can interpret the tortoise as lying about accepting what is written down, or not really understanding what he says he accepts. But in either case, and unlike the situation as presented by Carroll, the interpretation of the tortoise as understanding and accepting everything that Achilles has written without having accepted the conclusion will no longer be tenable, even *prima facie*.

This lesson is in a sense a lesson about a class of statements: imperative-mood sentential expressions expressing logical requirements. In particular, it is a lesson about how the addition of this class of statements to a language can *enrich its expressive power*: what the addition allows to be said, using these expressions, that could not have been said previously. We can think of introduction of such statements as providing users of the language with a way of leveraging linguistic competence into other kinds of competence. Once they are both competent in the use of imperatives, but not before, Achilles and the tortoise will be able to use this linguistic competence so that Achilles may communicate to the tortoise the basis of *practical* competence in reasoning as logic. This role of an expression-type – that of transmuting aspects of non-linguistic abilities to *do* into abilities to *say*, thus making them capable of linguistic transmission or expression – is, to borrow a term from Robert Brandom (1994, 2000) a job of *explicitating* something.

Consider Brandom's explanation of the explicating function of conditionals: 'The context "If... then..." allows us to express, in the form of a claim, command of the relations between the circumstances in which a claim is applicable and the consequences of its applicability. As Brandom puts it (2000, 60):

prior to the introduction of such a conditional locution, one could *do* something, one could treat a judgment as having a certain content (implicitly attribute that content to it) by endorsing various inferences involving it and rejecting others. After conditional locutions have been introduced, one can say, as part of the content of a claim... *that* a certain inference is acceptable.

The way that Brandom couches this explanation involves his view that, in every case, an inferential relation in which a claim stands redounds to that claim's content. But that commitment is detachable from the broader point here. Without a conditional locution, one might have an ability to *make* good inferences; but the conditional locution is required for *saying* which inferences are good. In this way, this fragment of vocabulary allows a bit of know-how to find expression as a knowing-that.

What I want to show is first: that expressions with believe(s)-that clauses may be used to perform an explicating function, and in particular that believe(s)-that clauses can be used to say what it is that reasoning compels of a reasoner. Second, my argument will aim to show that in order to perform this function, we must use this tract of belief-talk in a way which commits us to TN. If that is right, then, as Manifest Truth-Normativity says, a notion of belief as subject to TN is built into the structure of the practice of reasoning. Beliefs, understood as the things that reasoning issues in, are to be talked about as conforming to Truth-Normativity. Either they really do so conform, or this is an unshakeable illusion.

3.2 – The Fable of Helen of Troy and the Hare

I will present a twist on this fable in which Helen of Troy – whose role is similar to that of Achilles – attempts to explain to a hare what reasoning requires of him. Like the tortoise, the hare will be in most respects an apparently rational and competent interlocutor, but he will fail to understand something important about how to reason. However, in my telling, the hare will be the one with the writing implements in hand, and the gap in the hare’s understanding – and the corresponding gap in what Helen thinks to say (or knows how to say, or thinks she is allowed to say) – will be different. The hare will be perfectly ready to acknowledge *that* in accepting what Helen asserts, he is rationally compelled, but he will not grasp (or pretend not to grasp), *what* his acceptance compels. The moral of the story, I will argue (in 3.3), is:

Moral (H&H): In order to convey by *saying* all that one needs to know in order to reason, we need to be able to make claims about beliefs, either explicitly stated or implicitly understood to be subject to Truth-Normativity.

Let’s join our heroes now:

Helen: God, we are told, is omnipotent.

Hare: Indeed he is – let me write that down.

Helen: Moreover, God is omnibenevolent, is he not?

Hare: Indeed he is – I’ll put that on the board as well.

Helen: So ours is the best of all possible worlds! Put that on the board.

Hare: No.

Helen: What? If it is true that God is omnipotent and omnibenevolent, then it mustn’t it be true that ours is the best of all possible worlds?

Hare: Yes, of course; if there were a better possibility than the one we are living, an omnipotent, omnibenevolent God would have created that one instead.

Helen: Exactly! So why not write down that this is the best of all possible worlds?

Hare: Well, my writing paw is sore, and maybe I could power through it, but my handwriting is so bad when my paw is tired that it would barely be legible anyway... and besides, I think I'm realizing that I'm allergic to chalk dust, and...

Helen: No, no, no, the writing down itself isn't the point, it's just... will you at least say it back to me, and mean it?

Hare: Well, I'm hesitant to do that because, there's something about hearing those syllables in my voice that would just give me the shivers... *ugh!*

Helen: No, again, I guess I'm not making the point clearly. It's not that exactly that I'm asking you to make a specific utterance. What I'm asking for is something that is, sort of, "behind" a sincere utterance, if you know what I mean... I don't know what to call it...

The idea, of course, is supposed to be that what is missing from Helen's vocabulary is a tool with which to tell the hare *what* it is that being shown that something is true requires of him, namely, that he believe it. If she could clarify to the hare that all she meant was that he must *believe* that ours is the best of all possible worlds, the hare would no longer be able to maintain the position of apparently understanding and accepting everything Helen says without cooperating in her desired deliberative outcome.

Consider what would happen if we enrich Helen's vocabulary with belief-ascriptions and let the dialogue take its course. Starting in the middle, it would perhaps go like this:

Helen: What? Don't you see that if it is true that God is omnipotent and omnibenevolent, then it must be true that ours is the best of all possible worlds?

Hare: Yes, of course; if there were a better possibility than the one we are living, an omnipotent, omnibenevolent God would have created that one instead.

Helen: Exactly! And since I have shown you that it is *true*, I have shown you that you must *believe* it.

This modified dialogue illustrates the explicating function of belief-talk I am interested in. In the next section, I will argue that in order to perform this function, beliefs must be treated as subject to TN.

3.3 – The Core Argument

Helen, I have claimed, needs a notion of belief (and a word designating that notion) in order to be able to *say* to the hare what it is that he isn't grasping good reasoning to require, *viz.* that it requires that he believe what she has shown to be true. I will argue that, if the "belief" Helen invokes is to play this explicating role, it must be understood to be subject to Truth-Normativity.

The easiest way to see that this is so, I think, is to consider a notion of "belief" subject to different constraints, and to see why this won't do. Suppose someone attempts to use this kind of belief-talk but thinks that it is insufficient, in order for it to be the case that one ought to believe P, for P merely to be true. That is, suppose they think that the "left-to-right" direction of TN is false. I will argue that the way of using belief-talk consistent with this conviction would preclude it from performing the relevant explicating function of saying what reason compels. If this is right, then the use of-belief talk to make explicit what reason compels does presuppose the left-to-right direction of TN, and so the practice of reasoning itself, in which the same conception of what reasoning compels is *implicitly* accepted, itself presupposes the left-to-right direction of TN.

There are various theses that are in fact commonly maintained about belief that we might impute to the users of this particular sense of "belief." We can try to imagine a user of the locution who thinks that it is not universally the case that if p is true one ought to believe it, but rather that it is only universally the case that if p is true *and one has sufficient evidence for it*, one ought to believe it. Or we could suppose a user of this kind of belief-talk to think that it is only universally the case that *if p is true and it would not frustrate one's aims to believe it*, then one ought to believe it. We could even

suppose that a user of the locution understood that it is only universally the case that if p is true *and one is considering whether p* , then one ought to believe it, and still make the point that such a construal won't do.

To see the problem, consider the fact that the dialogue between Helen and the Hare has two parts. In the first part, Helen presents a line of reasoning meant to compel the hare. This part concludes when Helen exclaims "So ours is the best of all possible worlds!" After that point, Helen is searching for a way to say *what* it is that her reasoning from the first part compels. The locution "belief," taken in the relevant sense, is supposed to answer this need of Helen's.

But suppose "belief" is used alongside the thought that in order to compel belief, a proposition must not only be true, but under consideration. If this were the notion suited to explicate what is required in reasoning, then the "proper unabridged" version of the first half of the dialogue should end like this:

...

Helen: Moreover, God is omnibenevolent, is he not?

Hare: Indeed he is – I'll put that on the board as well.

Helen: *And what is more, it is under consideration whether ours is the best of all possible worlds.*

Hare: Correct – I'll write that down, too.

Helen: So ours it the best of all possible worlds!

But this edition of the dialogue does *not* portray Helen as making her argument more complete. Rather, it has her raising an irrelevant consideration.

It will not help to suggest that the considered-ness of the conclusion is not functioning here as a premise, but rather as a presupposition in the dialogue, and that the requirement on belief still

depends on it. This won't do because, although raising a presupposition on which the force of an argument rests may be redundant or otiose, it is not irrelevant. But when Helen raises the considered-ness of the conclusion here, it *is* irrelevant to the forcefulness of her argument.

The notion of belief as compelled only by truth in conjunction with considered-ness may be good for something. But it is no good for saying what we are compelled to do in reasoning.

There is one alternative that could go in as a putative necessary ingredient over and beyond truth if belief is to be compelled. This is the idea that a proposition's truth combined with *your having been shown that it is true*, suffices to compel belief. The worry is that if Helen says "and what's more, *I have shown you* that ours is the best of all possible worlds" this sounds much less like an irrelevant digression than talk of what is under consideration or what would help the Hare accomplish his goals does.³

Nonetheless, I think that with a little further scrutiny, we can see Helen's claim about what she has just shown the Hare is not functioning as a collateral element of what compels him to believe it. Consider what happens if we suppose the Hare to deny this, and tell Helen "No, you didn't show me, you showed someone else," or "no, you were just *about* to show me that, but hadn't started," but to remains aware and accepting of what Helen supposedly showed someone else or was about to show him. In such a case, it is still just as nonsensical as in for the Hare not to believe the conclusion. Nothing about who was shown by Helen, or when they were shown, is part of what should compel the Hare to believe in this instance. Based on these considerations, I think a better read of what Helen's insistence that she has shown the Hare that the proposition in question is true

³ Thanks to Asa Zabarsky and Susanna Rinard for pressing me on this point in an M&E workshop presentation of this material.

contributes is that it is just a way of emphatically drawing the Hare's attention back to the fact that the proposition must be true, since it can be shown, since it has been shown by her to him.

So far, the argument I have made supports only *half* of MTN. I have argued that we must understand our reasoning as involving just one direction of the bi-directional Truth-Normativity principle, namely the "right-to-left" or "if P, then one ought to believe P" direction. So it is still open that the notion of belief implicit, through the activity of reasoning, in the manifest character of our intellectual lives, is one on which each truth is to be believed, as well as some untruths besides.

But this is implausible. Indeed, we could make a parallel argument in the opposite direction: just imagine now that Helen is trying to talk the hare *out* of the idea that ours is the best of all possible worlds. Then in order for Helen to make explicit what her reasoning compels the hare to *refrain from*, she must cite a notion of belief that – likewise, because the reasoning that so compels should only include elements relevant to showing the proposition to be *untrue* – she understands as subject to the other half of TN: one ought to believe something *only if* it is true.

§4 – Elaborating the Manifest Normative Character of Belief: Objections

Replies

I want to consider three objections to this argument. I choose these objections for two reasons: first, these seem to me to be three of the most likely objections to occur, and second, answering them will help to show more of the contours of the notion of belief the necessity of which the argument above supports. In particular, in 4.2, 4.3, and 4.4, I will discuss an analogy between the use to which conceptions of belief and truth can be put in reasoning and the use to which game vocabulary and the corresponding conceptions can be put in winnable games. This

analogy will continue to be important in the following chapters as I advance and defend a notion of belief that is both plausibly attributable to us and plausibly governed by TN.

4.1 – Pragmatic Questions

Perhaps the most immediate worry about the argument I have given above is that it seems like we *can* make sense of “pragmatic” considerations favoring or opposing having a given belief, considerations that have to do with the benefits or burdens that would likely attend having this belief. Considerations of these benefits or burdens seem to be *orthogonal* to those of truth and falsity: it can be beneficial to hold a false belief, or burdensome to have a true one. For example, a runner who knows that she performs best when she expects to win, irrespective of whether this expectation is reasonable, may sensibly conclude that it would be best for her not to believe the truth about her competitors’ personal bests so she can go on expecting to win even if they have performed much better than her in the past. This case, and an unlimited number of others like it, one might worry, appear to make perfect sense of the idea that a proposition is true but should not be believed.

I want to suggest that there is another way of taking these cases. The argument that this is the *right* way to take these cases will have to await the elaboration of the positive conception of belief in chapter 3. But showing that will be necessary only to vindicate Truth-Normativity itself. For now, we don’t need to know whether this is the right way to think about the bearing of pragmatic considerations. Rather, we need to know that it is an *available* way to think about this bearing. In knowing this, we allay the worry that pragmatic considerations *force* us to think of belief as subject to normative constraints orthogonal to truth, which would undermine Manifest Truth-Normativity.

Consider the structure of the following case. You are standing outside of your house, and because it is chilly, you have prudential reason to be inside. But there is an angry porcupine blocking your door, and it is likely that if you try to go inside, you will end up inside-and-stuck-full-of-quills,

which, although a way of being inside, would be worse for you than remaining outside would be. The right tactic here seems to be to remain outside for the time being, while waiting for the porcupine to move. Your prudential reason to be inside, then, should not prompt you to pursue being inside in the manner that you typically would; you shouldn't walk towards the door just yet. But it should nonetheless constrain your behavior by preventing you from doing things that would cause you to lose your chance at getting inside. For instance, you shouldn't wander away and lose your chance to get inside as soon as the porcupine clears out. In this sense, you will be doing what you ought to only if you are treating being inside as a goal, and thus the claim "you ought to be inside" will correctly reflect the aims by which your behavior should be constrained. The presence of the porcupine provides a reason not to pursue being inside in the most direct way, and instead to pursue it in a way that is for now passive. It doesn't provide a reason not to be inside.

I am suggesting that we can understand the situation of the nervous runner, and in general, cases in which we can foresee that a subject's believing the truth would have bad consequences, along these lines. The runner's anxiety is the porcupine blocking the door to the truth. We can foresee that if she were to directly pursue the truth about her opponents' best times, this would have bad consequences for her. But it does not follow that it is impossible for her to have the belief without suffering the consequences. Rather, it remains possible for circumstances to change – she might visit a good sports psychologist, or she might even simply experience a spontaneous change in outlook – in a way that getting the truth and avoiding its damaging effects would be tractable.

In such a case, it may be that she should pursue a course of action that will most likely lead to her remaining ignorant for now, but which will keep her prepared, if circumstances change, to come to believe and make use of the truth. If this is right, it will still be the case that she ought to believe the truth, just not that she ought to believe it and be anxious about it, nor should actively

pursue it, since the latter outcome would be most likely for the time being. I will argue in chapter 3 that we can in fact understand all cases of potentially damaging true belief in this way.

4.2 – Compelling and Reminding

Here is another kind of objection. Consider that, in trying to persuade the tortoise, Helen might say “remember, we’re only concerned here about whether it really *is* the best of all possible worlds.” Or “remember we’re not asking about what would be *useful* to believe here, but just about what is the case.” In doing so, she doesn’t seem to be citing anything irrelevant. So why should we not consider the condition *that one is reasoning*, where reasoning is a specific kind of activity in which one is only concerned with the truth, to be an additional condition that must supplement the truth of a proposition in order for it to be the case that one ought to believe it (as far as the self-conception implicit in reasoning is concerned)?

We’ll consider the following analogy. Suppose you are playing a practice game of chess with your chess coach. But suppose that as a training exercise, she is having you play “king of the hill” chess, in which the first player to *either* checkmate the other *or* move their king to one of the four center squares of the board wins. Now suppose you are in a situation in which one move, m, would be best for traditional chess, but another move, n, would be best in king of the hill chess. Now suppose that, in a moment of absentmindedness, you begin to make move m. Your chess coach laughs as she sees what you’re doing and says, “this was a king-of-the-hill, game, not a regular game.” This stops you in the process of making m, and you proceed to make n without further ado.

Should we understand what your coach said as having had the effect of giving you a reason, in virtue of which you were compelled to pick n? I claim not. Begin with the assumption that there are just two ways in which you can make a claim that gives someone playing a goal-directed game like chess or king-of-the-hill chess compelling reason to do something. Either your reason is *game-*

internal; it comes from the goal-directed perspective one adopts in “playing to win” on which it is taken for granted that the aim of the game is to-be-pursued. Or your reason is *game-external*, in which case this is not taken for granted.

Your coach’s admonition cannot be giving a game-internal reason; from the point of view internal to the game, the fact that n is the better king-of-the-hill chess move is dispositive on its own. But it cannot be giving a game-external reason either; from the point of view external to the game, the fact that you are playing king-of-the-hill chess and is a better king-of-the-hill chess move are not by themselves compelling. From that perspective, there is an open question about whether one should keep playing one’s game of king-of-the-hill chess to win, or quit the game, or begin losing on purpose.

From the external point of view, what the coach says does not compel because it is insufficient. From the game-internal point of view it does not compel because it is otiose as a putative reason. But since in this situation, you *do* take yourself to be compelled to make n, we should understand what the coach says from the game-internal point of view. What she says functions not to add an independent reason, but merely as a *reminder* of the normative structure that is already *presupposed* to prevail on his activity by virtue of the goal-directed structure of his game-playing activity. You already knew you were committed to this structure – you simply failed to call it to mind at a key moment.

Similarly, we should think of the function of telling a reasoner that they are reasoning not as that of an internal reason (it would be otiose), or of an external reason (it would be insufficient) but rather a reminder, from the internal perspective, of what she is doing, and hence of the norms that this activity’s goal-directed structure presupposes.

4.3 – No External Perspective on Reasoning

But when we point this out, it leads to an additional objection. From the internal perspective, superiority of a move as a king-of-the-hill chess move compels taking it, but from the external perspective, one also needs a reason to continue playing king-of-the-hill chess to win. But if the analogy is to hold, then truth only suffices to compel belief from the perspective *internal* to the “game of reasoning.” But then why not adopt the perspective *external* to the game of reasoning in which one is engaged and ask whether there is sufficient reason to keep “playing this reasoning game” in the first place? Don’t we need some additional reason of *that* kind to collaborate with the truth, in order to give us compelling reason?

I accept the analogy. The perspective from which truth compels belief is analogous to the internal perspective in a winnable game. However, the objection does not succeed in showing that we require additional reason in order to be compelled to believe truths because there *is* no external perspective from which to regard reasoning. The bouts of reasoning in which one engages are not discrete separate games in the way that two consecutive games of chess are discrete token games. Rather, all of one’s reasoning is part of one long “round of the game.” The results of one’s previous bouts of reasoning make a difference to how to interpret each successive bout as successful or failed in a way for which there is no analogy in chess matches.

I’ll elaborate on this distinction a little more. First, notice that you don’t need to be occurrently doing anything chess-related (not even thinking about chess!) to be playing chess. Correspondence chess, in which you and your partner send one another your moves through snail-mail, makes this particularly clear. You might get the mail and see that your opponent, Windom Earle, has made a move that puts you in check, and then go to sleep. While you are asleep, you are in check in this game, and if you are in check, then you’re playing chess. Moreover, you still likely occupy the game-internal perspective. In this case, your having this perspective is a matter of your

propensities: what you would do if you were woken up, given time to become oriented, and had your attention directed toward the game. Although you are not doing anything that affects the state of your game with Windom, as long as you have a propensity to treat anything that *would* affect the state of your game from the game-internal perspective, you occupy the game-internal perspective in the relevant sense.

This is not to say that you cannot take the game-external perspective on your game with Windom once it begins. When you are discussing with your friend whether you should quit your game with Windom, you are adopting a perspective external to the game because it is an open question, from your perspective, whether you will subsequently do anything to continue the game, i.e. anything that has an effect on the state of the game besides to end it. What we need to see is why you should not be similarly regarded as adopting an external perspective when you wonder “is it really worth my time to reason about whether this is the best of all possible worlds in the first place?”

The first point is that you are still reasoning if you are wondering this. That is not enough, however, as the following elaboration shows. Suppose you and Windom are playing two simultaneous games of correspondence chess. To keep them separate, you write down your moves for one game in blue ink and the moves for the other in green ink. Then, while you are working on gaming out moves for the blue game, it might occur to you to throw the green game, since you are, for whatever reason, more invested in the blue game, and you’d like to reserve your chess stamina for it. Then you’d still be playing chess, and indeed actively exercising the internal perspective on the blue game, but you would be exercising the external perspective on the green game. Why can’t your stepping back from the question of whether this is the best of all possible worlds, to think instead

about whether it is worth your time to consider whether this is the best of all possible worlds, be like that?

To see why not, let's add just one more twist to your games with Windom. Suppose that instead of playing two chess games, you are playing one game of *Dual Chess*, where players play two "component games" at once and whenever a player *either*

- a) Puts the other player in check at move n for the same n in both component games, or
- b) Achieves checkmate at move m ,

One feature of dual chess that is usefully illustrative for my purposes is that there is no requirement, for the game to work, that you be at the same place in both component games at any given time.

You might be at move 15 in the blue component game and move 54 in the green component game; this is a perfectly valid stage in dual chess.

The fact that these games are linked means that you cannot coherently adopt the internal perspective on the game while adopting the external perspective on one component game, even if that component isn't your focus. For example, you cannot be focused on strategy in the blue component game from the perspective internal to dual chess while also wondering whether to throw the green component game. Throwing the green component game is throwing the whole game. You can perfectly coherently consider whether to (or even decide to) hold off on making any more moves in the green component game until the blue component game develops more. But in doing this you are still adopting the same internal perspective that you would be had you continued focusing on the green component game.

Moreover, because the course that one component takes makes a difference to the significance of the moves in the other component, you cannot properly "hive off" the perspective internal to a single component game and to treat it as a game of ordinary chess. If you are at move

15 in the blue component game, you ignore the fact that Windom had you in check at move 18 of the green component game at your own peril. To be sure, you might for strategic purposes, say to yourself “let me just focus on the blue game: can I force checkmate by move 19?” and game out your strategy from that point of view. But you cannot, from the game-internal perspective, be prepared to follow a sequence that would achieve checkmate by move 19 if it gave Windom the opportunity to put you in check at move 18 and win the game.

Similarly, you can “step back” from reasoning about one question to focus on another. And you might even *stipulate* an answer to your first question to help you in reasoning directed at your second question. But you cannot, from the perspective internal to reasoning, draw conclusions about the second question without heed to the consistency of these conclusions with conclusions you have drawn about the first question. You might be prepared to *give* up the conclusions from the first bout of reasoning if they conflict with those you get in your current one, but that’s a quite different matter from being *indifferent* to inconsistencies.

Go back to our original pair of questions for concreteness. Suppose you are now wondering whether it is the best use of your time to actively reason about whether this is the best of all possible worlds. It is one thing to decide, for the purposes of considering this prudential question, not to actively consider the metaphysical question, because you don’t expect doing the latter to help you in the former. But it is quite another to decide ahead of time to pay no heed to an argument if it establishes an answer to the prudential question by establishing an answer to the metaphysical question. If you are really reasoning about the prudential question, you are prepared to accept any argument that may be given or spontaneously occur to you if it shows you the truth with respect to that question (which is not, of course, to say that in reasoning on a question you are prepared to accept any argument that *purports* to show you the truth).

Because *all* of your episodes of reasoning are mutually constraining in this way, the internal perspective you exhibit in one bout of reasoning is the same internal perspective you exhibit in any other. There is not, strictly speaking, the game-internal perspective on reasoning about a particular question; there is just the game-internal perspective on reasoning.

Insofar as you have *any* propensities to engage in reasoning with regard to *any* topic, then, you are passively exercising the game-internal perspective on reasoning. That is why I say that you cannot occupy a game-external perspective with respect to the activity of reasoning, hence why the question “why go on reasoning?” makes no sense.

4.4 – Explicitating, Winnable Games, and Reasoning

I spent much of subsections 4.2 and 4.3 stressing the analogy between reasoning and winnable games. I have suggested that one aspect of the use of the notion of belief – the aspect my examination of belief is cleaving to – is its explicitating function. One thing that possessing a concept of belief is good for is that it enables us to *say* or explicitly think what, without the concept, we would know in the sense of know-how, but would be unable to say, about what reason compels. The concept that allows us to do that is the concept of belief, in my sense.

Now I want to connect these topics by discussing the analogy between, on the one hand, the explicitating role of “belief” with respect to reasoning and, on the other, the explicitating role of game vocabulary with respect to winnable games. This discussion anticipates the approach I will take in chapter 3 to understanding the “oughts” of Truth-Normativity in a naturalistically acceptable way.

It usually doesn’t happen this way, but one can perfectly well imagine that someone learns to play chess by carefully observing experienced players playing in silence, without having learned any chess vocabulary. Such a person may not be able to articulate a complete set of rules. (Perhaps you think that someone who has the ability to play chess must be able to do this in principle. I doubt

even this, but it in any case, it would take some sustained reflection to formulate them from scratch, and we can imagine our subject not to have engaged in this reflection). Such a chess player might in principle be as capable a player as any other, but she will be missing a tool in her chess-playing arsenal. Without the wherewithal to think “I’m in x position” or “that was a y opening” or “I’ll make move z,” she will play entirely by feel. It may well be that human chess players already do this most of the time. But an ordinary chess expert need not play entirely by feel. They also have the ability to think or talk out what has occurred in the game so far, what they should do next, what their opponent might do, and so on. And they can use this ability to supplement their ability to play by feel, or even to catch mistakes that they would have made were they only playing by feel. The “game vocabulary” (or concept-repertoire) plays the role of a tool that regulates and supplements performance. It does this by turning what one might know implicitly (as embodied in an ability to play by feel) – namely, such facts as “from position x, sequence y forces checkmate” into something that can be said explicitly.

We can think about the vocabulary and conceptual repertoire involved in thinking and talking about reasoning as playing an analogous role. A person can reason without thinking or talking explicitly about belief, truth, implication, etc. But if they are equipped with the ability to talk about such things, they have an additional method for producing positive results in reasoning and for catching mistakes, *viz.* thinking explicitly about reasoning, its aims, and its demands.

Now consider the different roles that different kinds of chess vocabulary can play. There are terms that describe the position of the board, either specifying the exact position, or specifying the position of some pieces, or otherwise narrowing down what the position is in the space of all possible chess board positions. To know, as a chess player, what it is to be in a position meeting such a description is not necessarily to know whether it would be a good position to be in. To

classify a position using this kind of vocabulary is not to commend or condemn it. Nonetheless, from the game-internal perspective, these kinds of classifications can be used to think and talk out one's play, in order to produce and regulate good play. But there are also classifications that are not neutral in this way. Consider the classification of a position as a checkmate. If you understand how to use this term from the game-internal perspective, then in classifying a position as a checkmate, you are commending it (if it is a checkmate of your opponent) or condemning it (if it is a checkmate from you). In the use internal to the game of the vocabulary to produce and regulate good play, the classification "checkmate" is *regulative* in the following sense: it is used not only to classify positions that count as checkmate, but to *produce* them.

Similarly, a characterization of someone as believing something is in itself neutral; to characterize a state as one of believing something is not, as such, to commend it or condemn it. But to characterize some proposition as *true*, I have argued, is to characterize it as to-be-believed, and to characterize a potential state of belief as one of true belief is to characterize it as of a kind which is to-be-pursued. "True," that is, like "checkmate" in chess, has a *regulative* role in reasoning. Talk of "true beliefs" can be used not only to classify true beliefs, but also in the process of producing them.

Chapter 2 – Teleology on the Heath

It stands to reason that the nature of the challenges which an attempt to vindicate TN faces is to a certain extent dependent on the approach such an attempt takes. The approach will focus on the construal of intellectual activity as the exercise of natural powers to achieve natural purposes. And I will construe belief's role in this purposively organized activity as the grounds of the requirements to which it is subject: for there to be requirements on belief in the relevant sense will be for beliefs to have a particular status *as* products of the exercise of this characteristic activity.

I want to clarify straight away what I mean in speaking of *natural* purposes. In talking of “natural” purposes, I do *not* mean to suggest a contrast with artifactual or elective purposes. Quite generally, although I recognize that using “natural” in opposition with “artifactual,” can be useful, I think it is ultimately misleading and this case is no exception, by my lights. Artifacts and stipulations are products of our thoughts and deeds, and therefore so are artifactual and stipulated purposes. And we should seek to understand our thinking and our doing as natural processes. Being that natural and artifactual purposes are the products of what we should seek to understand as natural processes, we should seek to understand these purposes themselves as natural, i.e. to understand artifactual and stipulative purposes as special cases of natural purposes. When I speak of “natural” purposes, I mean to delineate the class of all purposes of which a naturalistic account may be given. In this chapter, I aim to show that purposes of artifacts and stipulated purposes are natural in this sense.

My choice to understand the normative character of belief as a matter of its being the product of the exercise of a natural power will ultimately be best defended by the success of the approach, if I am right that it succeeds. But before diving into an extended discussion of the difficulties facing an approach of this kind, I will briefly discuss three reasons to think that it is an appealing approach to the problem I am setting for myself, given what I have said so far.

First, I have emphasized that in reasoning, and *unlike* in typical cases of playing a winnable game, there is apparently no way of embodying a “game-external” perspective. In this sense, the game of reasoning is “inescapable.” One thing that is suggestive about the natural purposes approach is that this inescapability bears a *prima facie* resemblance to the relationship between a living thing and the goals characteristic of its form of life. In evaluating how an individual plant is thriving, the “external” question of whether it would be better to find some other way of obtaining energy besides photosynthesis is moot. In being a plant, in having a life that is a plant’s life, it is already “baked in” that the plant will pursue a strategy involving photosynthesis. A question about whether it makes good sense to go on reasoning may seem similarly moot as a question about a human life.

Second, the notion of a natural purpose has been thought promising for the purpose of naturalizing the goal-directedness of apparently goal-directed activity. The conception of belief I am examining, I argued in the last chapter, must be understood as characteristically normatively constrained by the goals of reasoning. So the conception, if naturalistically plausible, needs to involve a naturalistically acceptable kind of purposiveness. In embracing the idea that this purposiveness is of a kind with the purposiveness of products of evolution more generally, I can draw on the work that has already been done to make purposes of this kind naturalistically intelligible.

Third, an attempt to do things in terms of natural purposes has the potential to afford me a unified approach to several questions my investigation has already raised. In this chapter, I will mostly discuss the idea that *requirements* of the relevant kind may be understood in terms of natural purposes. But in the next chapter, I will pick up on the idea, developed by various authors (including Karen Neander (2017), David Papineau (1993), and as most directly inspires me, Millikan), that *beliefs* may be understood in terms of natural purposes as well. And that’s not all: in chapter 1 I invoked

the notion of a *characteristic form of life*, and also that concepts and pieces of vocabulary can be essentially *for* something or other. It is a virtue of the natural purposes approach that rather than relying on separate tricks or assumptions to attempt to explain each of these four notions – requirements, beliefs, forms of life, and characteristic purposes of words and concepts – I can seek to account for each of them with a unified approach. The more I can explain in terms of the same tricks and assumptions, the fewer tricks and assumptions I need.

In what follows, I elaborate, and defend what I take to be a liberal, naturalistic, and realistic account of the commonsense notion of something's having a purpose or function in the teleological sense.

§1 – Optimism about Commonsense Teleology

There is a line of thought on which the biological, social, and artifactual aspects of the world are rife with functions in the teleological sense, that is, purposes. An antelope's heart is for circulating its blood; a natural language is for its speakers to communicate with; a watch is for keeping track of time. In a certain frame of mind, this line of thought has the ring of common sense. But it is another question – the question this chapter contemplates – whether it is a variety of the *best informed* common sense, a way of thinking that comports with our best scientific theories.

Reactions to this question vary. From one widely endorsed, *pessimistic* perspective, the idea of teleology in biology is to be repudiated as out-and-out mistaken: Darwin debunked the idea that the heart is for circulating blood. On another, *optimistic* perspective, far from debunking the common tendency to attribute purposes to aspects of the biological, social, and artifactual world, the theory of evolution by natural selection provides the framework for *vindicating* this tendency in an empirically respectable way: the notion of natural selection provides as a naturalistic rendering of what it means to say that the heart is for circulating blood, one which might be generalized into a naturalistic

rendering of what it means to attribute functions to learned behaviors, inventions, and the like. This optimistic perspective, though perhaps minoritarian, enjoys a distinguished pedigree, one which traces to Darwin himself (F. Darwin, 1887, quoted in Lennox, 1993).¹

Differences between optimists and pessimists about the fate of ordinary teleological thinking after Darwin undoubtedly spring in part from the fact that there is no unified “doctrine” of common sense. Different teleological claims or lines of thought may seem commonsensical in different cultures, at different moments in intellectual history, among different individuals, and even for the same individual at different moments or in different circumstances. Commonsensical ideas are usually at least somewhat ambiguous, and they are often wrong, to say nothing of their mutual inconsistency with one another. And true to form, at least some ideas about purposes in the biological world which might have seemed commonsensical to some are more-or-less directly refuted by our understanding of biological evolution. To the extent that the ideas that biological organisms are intelligently designed, or that organs “strive toward” or “foresee” a certain kind of functioning might be considered commonsensical ideas about teleology, they are commonsensical *errors* about teleology. Pessimists about commonsense teleology tend to focus on these sorts of ideas in maintaining their pessimism: evolution by natural selection explains the resemblance between biological entities and products of our intelligent design. But because biological evolution is a mindless process, it explains *away* the appearance that the likes of organs or behaviors possess anything akin to the purposes with which we imbue the likes of tools, rules, or ceremonies when we intelligently design them.

¹ Karen Neander (1991) and Daniel Dennett (1995), for example, contribute seminal modern defenses of this idea. Noteworthy recent defenses come from, for example, David Haig (2020 ch. 1) in biology and Justin Garson (2019, forthcoming) in philosophy.

Optimists agree, of course, that a notion of biological teleology that imputed will or foresight – whether to a supposed designer of organisms or to the activity of parts of organisms – wouldn't pass empirical muster. But they disagree with the suggestion that commonsense teleological thinking is folly through-and-through. Optimists think that if we consider the whole range of ordinary tendencies to think teleologically, we will find an admixture of insight and error, one from which a substantial kernel of truth may be extracted. They urge us to focus on the way in which products of biological evolution and of design appear to share a specific, unusual relationship between *explanatory structure* and *temporal sequence*. Some configurational features of a mousetrap *now* are by the link between these features and a possible *future* result – the trapping of mice – to which these features conduce. Similarly, various features of a fetal calf's eye *now* are explained by the effect on populations from which it descended of natural selection for the ability to *see*, something that the eye may or may not eventually do *later*. The theory of natural selection, then, might point us in the direction of a scientifically respectable notion of *final-causality*, to form a kernel of truth that remains in ordinary teleological thinking once we cast aside any naturalistically illicit notions of mentality or fore-ordination.

In this chapter, I aim to support a stronger optimism than has usually been thought possible of late. Roughly, it is the view that, across the biological, the social, and the artifactual, *things generally really are for what, when examined carefully, they seem to be for*. To do so, I will propose a new theory of functions and argue that the notion of function it introduces is suited to vindicate a substantial kernel of truth in a certain kind of ordinary thinking about purposes in the biological, artifactual, and social realms, and to do so in a naturalistic, empirically chaste fashion.

In section 1, I elaborate the strand of ordinary teleological thinking the correctness and empirical chastity of which I want to defend. I identify this strand with a commitment to the

legitimacy of a notion of function with three features and explain why I think such a notion is worth trying to work out. In section 2, I argue that no existing theory of functions supplies a notion of function of which the three features hold. Suitable as the various proposed notions of function may be to other contexts, none can capture the notion of function employed in the fragment of ordinary teleological thinking I aim to vindicate.

Section 3 presents a new theory of functions – the *Comparative Conduciveness Etiological Theory* – and discusses its principal innovations. CCET makes a new proposal about what processes of natural selection, design, and learning share in common such that their operation results in features of reality with final-causal explanatory structures. The proposal, in brief, is that these processes are all “selection-like” in virtue of the explanatory role within each processes of a *comparison between alternative possible features*, as to their *conduciveness to a certain outcome*. In section 4, I argue that identifying this explanatory structure common to the three processes jointly vindicates two of the *desiderata* on the notion of function I am after, features I call *Unity* and *Final-Causality*.

Further, CCET makes a novel distinction in the logic of function attributions, between a *functionary*, o, which is meant to perform some function, and a *beneficiary*, X which is shaped by a selection-like process to conduce to o’s performing that function. This distinction has the effect, in short, of allowing that an item (organ, behavior, artifact etc.) o may have the function of doing π if o’s doing π is a final cause of *anything*. This allowance contrasts with the requirement, common to the closest kin of my proposal, Selected Effects Theories, that doing π be a final cause of o itself or of some of o’s traits if π is to be a purpose of o. In section 5, I argue that this difference substantially liberalizes the notion of function and that CCET thereby satisfies the third feature of a notion of function I’m aiming at, one which I call *Appearance-as-Guide*.

§2 – A Variety of Commonsense Teleology

2.1 – Paley after Darwin: Refuted or Reformed?

We can think of the optimistic line on commonsense teleology I will defend as an optimistic line on William Paley’s “watchmaker analogy” from the opening passages of his *Natural Theology* (1963/1802). I present the key points of the thought experiment in the form of a fable.

A Walk on the Heath: One morning, the Reverend Paley went for a walk on a heath. There he encountered a stone, a watch, and an eyeball (of, say, a lizard). One of these three was not like the others. Examining the stone lent little insight into how it originated. He expected he’d even have a hard time proving it hadn’t simply been there forever (3). The watch’s structural arrangement, by contrast, seemed extraordinarily particular: it was arranged such that it could do something – tick with the passing seconds – that nearly every other possible arrangement of these same parts *couldn’t* do: “if the different parts had been differently shaped from what they are, or placed in any other manner... either no motion at all would have carried on in the machine, or none which answers the use that is now served by it” (*Ibid.*). This was excellent evidence, Paley thought, of its having been intelligently designed to keep track of time. The eyeball’s arrangement was special in something like the same sense: the laws of optics dictated that almost any other arrangement would not have allowed a lizard to see with it, but that the present arrangement would (13). This was excellent evidence that the eyeball intelligently designed for the lizard to see with.

Call the pattern of inference I’ve depicted Paley as endorsing the Heath Inference:

Heath Inference: The arrangement of things has the distinct appearance that σ is for doing π ; we may therefore conclude that it is intelligently designed to do π .

If we focus only on the conclusion, we can say that Darwin shows that Paley was flatly incorrect. The most famous instance of this common treatment of Paley occurs in Richard Dawkins’s book *The Blind Watchmaker* (1986). Dawkins’s tone towards Paley is generous. But all the same, his conclusion is, plain and simple, that Paley was “gloriously and utterly wrong” and that “the analogy between telescope and eye, between watch and living organism, is false” (5).

But there is being incorrect because you're pursuing a fundamentally wrong idea, and then there's being incorrect because you've taken a good idea too far. The latter kind of incorrectness should occasion reform, not repudiation. The *appearance* that watches and eyes are importantly analogous already struck Paley as quite strong. And, at least as *appearance*, it has only gotten massively stronger. You want to talk about parts which "had [they] been differently shaped... or placed in any other manner... wouldn't [function as they do]"? Reverend, just *wait* until you hear about proteins! Perhaps, then, Paley's thought experiment is a fundamentally right-headed way of identifying *that* the biological and artifactual share something important in common, and his error was to misidentify *what* this something was.

The indication we find in evolutionary theory of how Earthly life developed from a common origin into the variety and complexity we see today may even positively *recommend* a line that largely conserves Paley's reasoning. For this discovery provides a perspective from which the biological and artifactual make up a *genealogical unity*. What I mean is this: there is a corner of the universe, arranged in a big, branching, spatiotemporally connected structure, and within and throughout this big branching process, patterns are exhibited the likes of which are not seen for light-years around. This big branching structure is called "the history of all known life." But notice that the history of all known artifacts is connected to this branching process. Artifacts emanate out of certain specific tips of the branching process of terrestrial life: human fingertips (or, sometimes, the extremities of the likes of birds, beavers, and bees). A repudiation of Paley's thinking would construe the similarities between the kind of order exhibited in biological pieces of this branching structure and that exhibited in the artifactual emanations from that structure as superficial or coincidental. But all known examples of organisms and artifacts issue from just one process: not one *type* of process, but one *token* process! Bearing this in mind, we might be almost irresistibly tempted to view the similarities we perceive as not shallow or coincidental, but as owing to a common origin in that process, a result of the

extension and ramification of what are, broadly, the same principles of order operating throughout it.²

Instead of abandoning the inference embodying Paley's line of thought, this perspective would pare back Paley's conclusion so that the inference is confined within proper bounds. Doing so yields the

Reformed Heath Inference: The arrangement of things has the distinct appearance that o is for doing π ; we may therefore conclude that o is *for* doing π (π is a function/purpose of o).

If we hope to preserve core features of Paley's perceived analogy between the artifactual and the biological while jettisoning his commitment to an intelligent designer of life, just what might we hope to preserve? I suggest the three features.

The first derives from the idea that Paley has really identified a substantive analogy between the evolved and the designed.³ The notion of function we apply in the conclusion of the RHI should apply univocally to each. Call this feature *Unity*. Note that Unity is compatible with the idea that biological and artifact functions may be importantly different species of a genus.

The second feature keys in on the almost backwards-causation-like structure of evolutionary and design explanations, in which the sort of thing that something may (or may not) *go on* to do figures in evolutionary and design explanations of how things came to be thus organized *now*.

² Dennett's "Tower of Generate and Test" (1995) is one noteworthy effort to work out this idea.

³ According to the official view here, the things between which the substantive analogy is supposed to obtain are products of (genetically or memetically mediated), *evolution*, of *intelligent design*, and of *learning* and their parts, traits, or aspects. As I do here and have already done above, I sometimes refer to this official subject matter in terms that don't quite exactly pick it out, as amenability to context dictates (for example here, because Paley's argument from analogy doesn't focus on products of learning). For instance, when I deploy such dyads as "the biological and the artifactual" or such triads as "the biological, the social, and the artifactual," I am referring, in an imprecise way, to the "official triad" of evolved, learned, and designed, meant to emphasize that biological, social, and artifactual phenomena are replete with products of evolution, learning, and design.

Focusing on this point, we may hope to preserve, and make naturalistically respectable, the idea that what Paley recognizes in both the watch and the eyeball is real *Final-Causality*.

Last is a property not exactly of functions themselves, but of the epistemology of function-ascriptions. We might hope that our reform leaves us with a notion of purpose on which the RHI is a warranted inference, one on which the kind of appearance of purposiveness Paley recognizes in both the watch and the eyeball is a reliable guide to reality about purposes, at least when we are considering products of evolution, learning, or design, or these products' parts or aspects. Call this property *Appearance-as-Guide*.

It bears emphasizing that Paley doesn't rely on any old shallow or un-scrutinized appearances to draw conclusions about things' purposes. He makes sure to scrupulously study the item's current workings, and draws on hard-won background knowledge in, for example, mechanics and optics. He relies on appearances to draw an inference only in this sense: when an appearance that a thing is purposive persists given an adequate understanding of its present-day workings, he infers that it actually has this purpose (pre-reform: that it was designed). I will try to vindicate appearance as a guide to reality about functions in something like this sense; I don't mean to suggest that superficial appearances or first guesses about a thing's purpose are particularly apt to be accurate.

2.2 – The Usefulness of Unity, Final-Causality, and Appearance-as-Guide

Above, I have tried to pump the intuition that there could be a salvageable analogy in Paley's thought experiment, embodied in a notion of function satisfying Unity, Final-Causality, and Appearance-as-Guide. To this, I will add a few comments on the potential usefulness of the sort of commonsense teleological thinking such a notion would enable. It is worth seeing whether thinking

this way about purposes can be legitimized, because if it is legitimate, it holds some promise to help us understand things better.

Consider the range of parts of our world which have been shaped by biological evolution (bacteria, elephants,...), intelligent design (can openers, websites,...), evolution through memetic transmission and selection (slang, accents...), or even through some difficult-to-identify combination of one or more of the three (political ideologies, codes of ethics, modes of production,...). Now consider the range of projects that hope to advance our understanding of some such phenomena by appealing to teleological concepts. Take, for instance, teleosemantics in the philosophy of mind, or constitutivist approaches in meta-ethics and epistemology, or even functional or ecological approaches in the social sciences and social criticism. These projects, if they are to be naturalistic, rather than non-naturalistic, if they are to be realist rather than anti-realist, and if they are to make confident assertions rather than mere speculations, rely on the legitimacy of a concept of function that applies in a univocal way to the designed and the evolved, is final-causal in character, and to which appearances, if they hold up under scrutiny, are a guide.

Unity is crucial for handling cases that suggest the RHI but are intermediaries between, mixtures of, or ambiguous between the evolved and the designed. Intelligent behavior might be a liminal case, sitting on the surface that forms the boundary between the evolved and the designed. Biological traits bred into a lineage by humans exhibit a kind of hybridity between the biological and artifactual (Sperber 2007). And many features of our social arrangements which seem manifestly purposively organized are mixed or ambiguous: are the norms of formal dining etiquette evolved or designed? What about the use of money? English Common Law? If we have univocal readings of teleological locutions, then, when faced with cases that we are tempted to place at the border between evolved and designed, that we are uncertain as to whether they are evolved or designed, or

that are likely partly evolved and partly designed, we may rightly remain undeterred in drawing inferences about what things are for.

Final-Causality is desirable if we want a notion of function to help clarify the place of practical or evaluative significance in nature, as for instance, a *eudaimonistic* meta-ethics or some versions of teleosemantics would aim to do. One way to see the connection is to think about the metaphors we find available in relation to teleology. When something is for doing π we might say that the doing of π is “inscribed” in a special way into the course of events that brought the current situation about, even if the item has not yet done π . A teleological perspective on something seems to attribute some real “current” carrying things toward the doing of π .⁴ The dynamics of a situation that these metaphors evoke seem especially well fitted to certain kinds of practical and evaluative questions. *Shall I go with the flow of the current carrying things toward the doing of π or shall I resist it? How does this thing measure up against the standard inscribed in it of the doing of π ?*

It probably goes without saying that it would be useful to have a notion of purpose of which Appearance-as-Guide is true. If it is possible, we would like a notion that we may comfortably ascribe on the basis of careful study of the items we find on the heath, combined only with the well-confirmed theories we have prior to embarking upon our walk. Without such a principle as Appearance-as-Guide, attempts to deploy teleological concepts in making sense of the world run up against baffles raised by the obscurity and complicatedness of actual evolutionary history. This is to say nothing of the extraordinary difficulty of corroborating the etiology of many social practice or institutions, which if required for before any teleological thinking about the social could take place, would make such thinking near impossible to attain warrant to do.

⁴ Obviously, we cannot really get clear about what final causes are unless we can get beyond metaphor. Part of my aim in proposing CCET is to take a stab at clarifying the explanatory structure characteristic of final causality.

It might seem that aiming for the conjunction of Final-Causality and Appearance-as-Guide is ill-fated. After all, the idea that evolution naturalizes final causes depends on the distinctive etiologies of items forged in the fires of biological competition. How, then, could such final causes be attributable without independently corroborating the relevant etiology? Well, consider this example:

Pedigree: You walk out your door and see two near-identical finchy-looking birds to your right and one much larger black bird to your left. You conclude that the two finchy birds share a nearer common ancestor than either does with the black bird.

Here, you draw a conclusion about the etiologies of three individuals. But it is a modest enough conclusion that you may safely, albeit fallibly, draw it without investigation into these individuals' histories. What we want is a notion of function that, though it has an etiological dimension, is undemanding enough that it regularly tees up inferences like the one in Pedigree.

§3 – Existing Theories of Function

No extant account of functions satisfies Unity, Final-Causality, and Appearance-as-Guide. I don't claim, of any of these notions of function, that they must therefore be regarded as illegitimate. But none can be the notion of function that underwrites the specific brand of commonsense teleology considered here.

The wide array of non-etiological theories, which do not place requirements on how some bit of functional organization came about, fail to be genuinely final-causal. A noteworthy early attempt at a genuinely final-causal etiological theory does not do the trick. Extant selected effects theories improve on previous etiological theories and furnish a notion that is genuinely final-causal, but they are ambiguous at best with respect to Unity and do not guarantee Appearance-as-Guide.

3.1 - Non-Etiological Theories

According to non-etiological theories of functions, a function attribution does not say anything in particular about the history of the item with the function itself. Versions of such an approach have included, for example, proposals to identify something's function with the way in which it contributes to fitness (Bigelow and Pargetter, 1987), the way in which it could contribute to fitness (Nanay 2010), the way its propensities contribute to a complex system of propensities (Cummins 1975), or the way in which its propensities contribute to the maintenance or persistence of a complex system (Moreno and Mossio 2015).

It is this distinguishing feature of non-etiological theories that renders them straightaway unsuited to vindicate the RHI. Such views ascribe functions to swamp-things, in whose origins nothing remotely like a final cause figured. They advance notions of function which are unabashedly not final-causal.⁵

Proponents of a notion of function as final cause need not quarrel with non-etiological theories of function. They can maintain that that a teleological, etiological, notion of function can peacefully coexist with a non-teleological, non-etiological one. One forceful voice urging this way of looking at things – distinguishing between an etilogically grounded, teleological notion of function, and notion of “how something functions” – comes from Millikan (2002).

One point raised against such a pluralistic attitude – for example, by Davies (2001) and recently by Andrew Rubner (2023) – is the supposed absence of an etiological notion of function from biological practice. Biologists do not engage in extended inquiries into a trait's evolutionary history in order to determine its function; they focus, rather, on what the trait presently does for its

⁵ Mossio and Bich (2017) argue that biological teleology can be understood in non-etiological, thermodynamic terms, as a matter of self-organization. It is not clear that this successfully captures final-causality; it appears incapable of assigning functions to defective items which do not as a matter of fact play the role they are “supposed to” in self-organization. Moreover, Mossio and Bich disclaim Unity, arguing that their theory draws out a difference between the teleology of artifacts and that of biological entities, and giving no indication of how we might understand them as species of a genus.

bearer. One explanation of this fact would be that biology only has room for a merely descriptive notion of function and that the teleological notion of function is to be discarded, or at least banished from biology. But here is another possibility: suppose that the RHI is justified because there is final-causal notion of function on which the etiological constraints are minimal and “teleology comes easily.” This would make the peaceful coexistence of etiological and non-etiological functions easier to recognize already in biological practice: the two notions would be much closer together in extension than has been generally assumed, rendering it mostly unnecessary for biologists to explicitly distinguish between them, or to engage in etiological inquiries in advance of attributing functions.

3.2 – The Simple Etiological Theory

The next category is a category of one, set off from others in this discussion because of its resemblance to my eventual proposal. It is a version of the Wright’s account (1973), slightly modified and simplified in ways that simplify the discussion here. The crucial clause of Wright’s definition requires that if o is for performing π , then it must exist in the form it does *because* it does π . Going by first appearances, there is hope that it is final-causal. But this hope is short-lived. The following example from Christopher Boorse (1976), makes the point decisively:

Boorse’s Knockout: A scientist is performing an experiment involving a hose conducting chlorine gas. A leak springs and knocks her unconscious, so she can’t repair it. Once the scientist has been unconscious for a time, it will be true that (the present time-slice of) the leak is around because (time-slices of) leaks in chlorine gas lines have a propensity to cause scientists to remain unconscious.

In this situation, the leak meets the requirements set by the Simple Etiological Theory to be *for* keeping the scientist unconscious. But scarcely anyone would endorse this verdict. The etiological condition of the simple etiological theory ascribes functions where there no teleology is to be found.

3.3 – Selected Effects Theories

Selected effects theories identify items' functions with something along the lines of: those effects the items were selected for having, elaborated differently depending on the particular theory. One way to understand the theories of this family is to see them as tightening up the etiological constraints of the simple etiological account to rule out the likes of Boorse's Knockout. Rather than requiring only that something be around because it can do π , selected effects theories require targets to have come about through *selection* for doing π . Hearts are around not just because they circulate blood, but due to the influence of *selection* for being able to circulate blood. The hole in the chlorine line that keeps the scientist unconscious is not a product of selection for being able to keep a scientist unconscious, so that is not its function.

3.3.1 – Selected Effects Theories and Unity

Some selected effects theories, such as that of Justin Garson (2017, 2019a) or selected-effects-adjacent proposals such as Marc Artiga's (2021), are explicitly intended to apply only to biological functions. They can therefore proffer notions of selection that straightforwardly do not satisfy Unity, having no need to do so.

Others hold that the way in which artifacts are made counts as a selection process in the relevant sense and take selected effects theories to capture a notion of function common to biological items and artifacts. Such proposals take on the burden of providing a sense of "selection" on which artifacts and biological entities alike are products of selection. Doing so has proven difficult. For instance, Paul Griffiths (1993) suggests that artifact design is a sort of selection process because it involves a designer generating several imagined prototypes, simulating their functioning, and eliminating those that don't perform as desired. Karen Neander (1991) observes that this psychological description seems much too specific to describe all cases of what we would call

intelligent design. Engineers often design things intelligently by calculating, rather than imagining what their design is to be. Sometimes we don't need to dream up a set of alternatives because a perfectly good design is obvious. Neander maintains that artifact functions are products of selection, but, having criticized Griffiths's attempt at spelling out what the relevant kind of selection amounts to, she declines to assay a new theory of selection in artifact design, leaving the notion unanalyzed. Millikan's selected effects theory (1984) is explicitly disjunctive, albeit along slightly different lines, dividing between entities which share their functions with lineages of ancestors, and entities, whether biological or artifactual, which derive their functions from having been produced by something which itself has a function.

The jury is out on whether a selected effects theory can deliver Unity. To do so, it must provide a univocal notion of selection encompassing both natural selection and artifact design.

3.3.2 – Selected Effects Theories Appearance-as-Guide: The Spandrel Problem

But confining our attention to biological cases, there is a further problem. Extant selected effects theories do not satisfy Appearance-as-Guide. All require that a trait emerge or be retained due to the influence of selection pressure to perform the function it is to be assigned. To suppose that every biologically useful trait emerged, or even that it has been retained, because of selection pressure to function as it has actually been useful is to hopelessly oversimplify how evolution works. The difficulty here is sometimes called the problem of *spandrels*, after the titular example of Gould and Lewontin's landmark essay (1979). In that essay, the authors ask us to consider the Cathedral of San Marco, whose dome is supported by arches, which arches intersect in smooth, concave triangular patterns called spandrels. These spandrels are decorated in a way that complements the overall iconography of the cathedral's interior. It's clear what they're supposed to be doing in the

overall design, namely depicting one evangelist and one river-personification each. Just ask the designers, and they'd tell you as much.

But although the spandrels perform this aesthetic function, it is not the reason they are there. The arches need to be there to hold up the dome, and the spandrels, the surfaces formed where the arches intersect, need to be there because arches do not tessellate. The spandrels didn't originate, nor do they stick around because of the purpose they serve. They are merely side effects which have been pressed into service. "Spandrel" has come to be used to classify biological traits as akin to the actual architectural phenomenon Gould and Lewontin deploy in their example. It is a contested matter what exactly should be meant by calling something a "spandrel" (Gould 1997). Here, I stipulate that a spandrel is a trait that is determinately useful in a certain way, thus that has a scrutiny-resilient appearance of purposiveness in the sense relevant to Appearance-as-Guide, but which is not there because it is useful in this way.

Extant selected effects theories don't satisfy Appearance-as-Guide because they deny spandrels their apparent functions. Spandrels are not there because of that which they seem to be for. And if that is your etiological requirement for function-attribution, then it is too demanding to allow these traits their apparent functions. Weakening the etiological condition in your notion of function by appealing to the effect of selective pressure in merely *retaining* a trait, as Peter Godfrey-Smith (1994), Millikan (1995), and Garson (2017, 2019) do, helps somewhat, but does not fully resolve the issue. Consider the Snowy Owl, *Bubo scandiacus*, whose young hatch exclusively above the arctic circle, and are feathery at birth, and begin to wander away from the nest weeks before being able to fly (Watson, 1957, König and Weick, 2008). Snowy owls' featheriness helps them fly and keeps them warm. But strictly speaking, it is just the role in thermoregulation that acts to keep the feathers around: a featherless snowy owl will die of cold before he dies of flightlessness, so the

conduciveness of feathers to flight never has the chance to play a part in selecting against featherlessness.⁶ Nonetheless, the featheriness of the Snowy Owl appears to be for flight.

Appealing, as Dennett does (1995) to the role of selection in optimizing a part does not take care of all cases, either. A trait may be used for both A and B, but optimized for B, and not A, because doing B is simply a much finickier matter. Consider the ‘specs’ blood must satisfy in order to circulate in the vascular system – it must be a not-too-viscous liquid – in comparison with the those it must satisfy to play its various other roles.

Because evolutionary histories are complex and opaque matters, extant selected effects theories do not legitimate Appearance-as-Guide.

§4 – The Comparative Conduciveness Etiological Theory

Here, I introduce a theory of functions which, I argue, is the first to jointly satisfy the three aspirations discussed above. It is univocal as concerns biological and artifactual purposes in much the same way that Wright’s Simple Etiological Theory is and therefore satisfies Unity; it is genuinely Final-Causal along much the same lines that selected effects theories manage to be; and it is undemanding enough that it has an extension very close to the extension of a typical non-etiological theory, thereby delivering Appearance-as-Guide.

I follow up the introduction with a brief discussion of the way in which the theory attributes functions to defective items incapable of performing those function, such as broken artifacts or malformed organs. The principal purpose of the discussion is to help the reader better understand the contours of the view by seeing how it handles cases that are in some way deviant. But it also

⁶ This argument applies only to the determinable trait of *featheriness*, not to any determinations thereof that the Snowy Owl exhibits. It would be a revision of most selected-effects theories to claim that a trait like featheriness could be for flying because a *determination thereof* is selected for flight (though not all; Millikan (1984) is a notable exception). In any case, this modification only suffices to handle some cases, as I discuss immediately below.

serves as an indication that the theory does indeed make the intuitive function-attributions to defective items, a feature which is commonly held as a *desideratum* of theories of functions (Garson 2015, 2019).

4.1 – Introducing the Theory

The first innovation of the theory is its strategy for jointly attaining Unity and Final-Causality. “Boorse’s Knockout” is a double *entendre*. It describes what happens to the scientist in Boorse’s counterexample. It also describes the effect that this refutation of Wright had on the whole approach of trying to spell out teleology in terms of “because.” As far as I can tell, when Wright’s version failed, the approach was dropped and not tried again. My theory revives this general approach to a theory of teleo-functions. It avoids Boorse’s Knockout by complicating the “because”-involving etiological clause in the *spirit* of a selected effects theory. Rather than requiring only that the current arrangement be explained by its being able to do π , I require that it be explained by its being *better* able to do π than some alternative. I will suggest that this etiological condition captures the spirit of the requirement of selected effects theories that *selection* for doing π , not just doing π , must figure in the explanation of how the present situation came about.⁷ But, in sticking to “because,” CCET achieves Unity without facing the burden of furnishing a univocal notion of selection.

The second innovation of CCET is its construal of function attributions as relating functions to not one but two items: there is the *functionary*, which is the thing that is meant to perform the function in question. And there is the *beneficiary*, which is whatever is organized around the

⁷ Actually, I believe a stronger claim: CCET is what you get when you take a notion of selection defined in terms of “because” and “plug it in” to the right sort of selected effects theory, so CCET *is* a selected effects theory that deploys definition of “selected” in becausal terms. If this is right, then CCET does not just evade, but solves the difficulty for selected effects theories about Unity. The argument for this suggestion must await another occasion.

functionary's performing its function. This distinction allows a substantial liberalization of the necessary conditions for having a function. Intuitively, the idea is that rather than an attribution of the function π to o requiring that o 's doing π be a final cause of o or of some characteristic of o , CCET requires merely that o 's doing π be a final cause of *something*. This liberalization allows for traits to have the functions they appear to have, even if they are spandrels, in virtue of the complementary organization of other elements of the unity of which the trait is a part.

Now, the proposal:

Comparative Conduciveness Etiological Theory (CCET): *With respect to X, o is for doing π just in case, for some properties F and F*, X is F rather than F* because X's being F is more conducive to such things as o doing π in some conditions C than would be its being F*.*

To sum up the main ideas: there is the object o , which I call the *functionary*, which is meant to do some π . There is the thing X, which I call the *beneficiary*, which is organized around the functionary's performing π . It is X, the beneficiary, not necessarily o , the functionary, which must have an etiology characteristic of purposive organization. The beneficiary must have some property, F – call this a *mark of purposive organization* – rather than some alternative, F*, because being this way is comparatively better for getting o to perform π – at least in some *normal conditions*, C (typically, these will be circumstances in which the item evolved or the circumstances in which the designer envisioned its being used).

For example: for Tweety (=X), Tweety's wings (=o) are for flying him around(= π). This purpose for his wings is constituted by such marks of purposive organization as the hollowness of his bones. He is hollow-boned (=F) rather than solid-boned (=F*) *because* being hollow-boned is more conducive to one's wings propelling one's body through the air, in terrestrial atmospheric and gravitational conditions (=C), than being solid-boned would be. Notice that these choices of mark

of purposive organization, alternative property, and normal conditions (F, F*, and C) are probably not the only ones which would bear out the function-attribution in question.

Separating out the functionary – the o meant to do some π from the beneficiary – the X organized around o’s doing π – means that functions are *relative* features of things: having a function is a matter of having that function in relation to some specified beneficiary. But this does not make having a function any less of an objective matter than being a cousin is. Moreover, this reading of the logic of function attributions is motivated, not *ad hoc*. Consider nature’s widespread antagonisms. An antelope’s heart is, with respect to the antelope, for circulating its blood, in virtue of the antelope’s adaptations which conduce to its doing so. But relative to the lion’s purposes, it is for being eaten. The lion possesses a suite of traits adapted to tracking down, killing and digesting such animals as antelopes, and she has these traits rather than others precisely because they are more conducive to success in making such use of such prey animals. Granted, we make sense of the talk of what the antelope’s heart is for *simpliciter*. But this can be understood as a reflection of a tacit, cancelable assumption that, when talking about a biological organ, we mean to talk about its functions with respect to its organism.

In a function attribution, the beneficiary X and functionary o may be identical, in which case we may call π a *proper function* of o. A well-built watch has proper functions: the watch itself exhibits its form rather than others because that form better conduces to the watch itself telling time. Organs whose forms are adapted thorough natural selection to the performance of their functions have proper functions as well.

X may be some whole of which o is a proper part, in which case we may call π an *organic function* of o. Though biological organs, behaviors, traits, and the like *can* have proper functions, I think that our commonsense thinking of functions in the biological world is generally more keyed

into the organic functions of these things with respect to the organisms of which they are a part. Thus, however Tweety's wings came about, it is an organic function, with respect to Tweety, that his wings fly him around because his bones are adapted to conduce to his wings doing so.

Finally, X may be separate from o, in which case we can say that π is an *improper function of o*. The sense in which the gazelle's heart is for the lion to eat is as an improper function, a function with respect to the lion. If, on a sunny day, I stand under a rock ledge for shade, it is an improper function of the ledge, with respect to me, to shade me: I am under it rather than a few feet over because being here is more conducive to the rock's shading me than being there would be.

4.2 – Functions for the Defective

A theory of teleo-functions should accommodate the idea that items, both biological and artifactual, may have functions which they cannot perform. It may not be obvious that the CCET does this. Aren't defective devices distinguished by a *lack* of organization tailored to their purposes? But: notice that the requirement in order for o to be for doing π is effectively, that o-doing- π -conduciveness is a final cause of *some* feature of *something*. This open-ended requirement results in many potentially redundant pathways to functionhood. There may, in a variety of circumstances, be marks of purposive organization for o-doing- π which ground the relevant function-attribution, even though these marks do not amount to features sufficient to allow o to do π . I'll survey a (non-exhaustive) sampling of routes to the attribution to defective items of their intuitive functions.

If we are dealing with a proper function of a complex functionary(/beneficiary) like a watch, the absence of some marks of purposive organization (e.g. of crucial springs) may render it defective while the presence of others (e.g. the usual placement of gears) render the ordinary function-attribution in order.

With organic functions, the functionary may have its intuitive functions even if its intrinsic features make it defective, because of the presence of marks of purposive organization elsewhere in the beneficiary. So for example, a malformed eye may be for seeing because of the marks of purposive organization present in the optic nerve and the visual cortex, i.e. the features thereof shaped by natural selection for cooperating with the eyes in order to thereby see.

Failed inventions may have their intended functions as described in the preceding two paragraphs. Or, when they are designed based on mistaken beliefs, they may have their intended functions with normal conditions which are counterfactual or counterpossible. For instance, suppose you make a y-shaped rod with the intention of using it for dowsing, a pseudoscientific practice which is supposed to locate ground-water. Your rod has a “y” shape rather than any other because that shape is conducive to locating groundwater in the counterfactual circumstance – which you mistakenly believed was the actual circumstance – that dowsing is a sound practice.

§5 – Comparative Conduciveness, Unity, and Final-Causality

CCET is not beholden to the search for a notion of selection that applies univocally to the processes that produce both evolved and designed items. Like Wright’s simple theory, it does things directly in terms of “because”. But unlike Wright’s theory, CCET’s focus on explanation by appeal to *comparative*, rather than absolute conduciveness to the performance of a function allows it to avoid attributing functions to elements of simple feedback loops like the leak in Boorse’s Knockout.

When a comparison between features according to their conduciveness-to-o-doing- π is explanatorily relevant to some X’s character, this means the following about X: it has been influenced by the workings of a process that

A) is sensitive to comparisons among alternative features with respect to their o-doing- π -conduciveness, in a way that

B) systematically favors the instantiations of more o-doing- π -conductive features.

Natural selection is sensitive in this way to comparisons between alternatives because it involves *competitions* between such alternatives. Intelligent design is sensitive to such comparisons between alternatives because it is conducted by minds that can think about these comparisons. Processes that are comparison-sensitive in this way “shape” or “funnel” or “pressure” X *toward* exhibiting those features which better conduce to o’s doing π . In cases like Boorse’s Knockout, on the other hand, *nothing* is shaped by *anything* that is systematically sensitive to *comparisons* between propensities to keep scientists unconscious, hence there is no feature born of a “pressure” to keep scientists unconscious, hence nothing organized around or “toward” this aim.

Below, I expand on the analogy CCET captures in products of evolution and intelligent design (and learning). Then I explain in more detail how neither Boorse’s Knockout nor a more threatening variant exhibits the kind comparison-sensitive etiology needed to count as purposive on CCET.

5.1 – Comparative Conduciveness Etiology in Natural Selection, Learning, and Intelligent Design

In the biological case, the notion of a fitness landscape is useful for conveying the role of comparison. Visualize a population of organisms as a cluster of points somewhere on a landscape of rolling hills. Each location on a landscape represents a possible combination of traits, and the elevation of each point corresponds to the degree of fitness such a combination confers. The theory of evolution by natural selection says that, over time, the clusters climb uphill on these landscapes. When a trait comes to predominate in a population because it newly allows its bearers to perform, or

better perform a given function – say, for example, the evolution of canine teeth to better tear meat – we can understand this as an effect of “evolutionary anti-gravity” pushing the population away from some less effective traits and towards other, more effective ones.

Consider polar bears’ white exteriors. In arctic climes, being white is *better* at causing you to go undetected than being brown. Thus, due to the history of competition between brown and white bears in arctic climes, it eventually came to pass that the population of polar bears was uniformly white, rather than brown. Hence, the previous generation of polar bears was exclusively white. Hence, that generation was apt to exclusively produce white offspring. Hence, Paul the polar bear, having been produced by such a population, will be white. Tracing the explanation back to where we started, Paul is white because being white is better for camouflage than being brown is.⁸

This comparative pattern of explanation fits intelligently designed items as well. A knife has a sharp edge rather than a dull one because a sharp edge is better for cutting. This is so because the knife’s maker knew that sharp edges are better than dull ones for cutting, so she made sure to make a knife with a sharp edge rather than a dull one. More generally, whatever exactly intelligent design

⁸ There are notorious worries about whether evolutionary explanations can explain individual traits such as Paul the polar bear’s white appearance. Forbearing an extended discussion of the relevant debates I’ll make a few brief points. Most arguments that evolution by natural selection cannot explain the traits of individuals are directed at purported patterns of explanation different from the one I have rehearsed just above, and which I do not think apply to it. Accepting the pattern of explanation I have proposed does not force us to reject genetic essentialism, nor to accept that evolution can explain the initial emergence of traits. To reject the validity of the pattern of explanation above, an objector would have to deny either.

1. That evolution can explain the fact that a less-than-uniformly white population became uniformly white, or
2. That Paul’s whiteness can be explained by the uniform whiteness of the population that produced him.

The kind of explanation in (1) is the canonical kind of explanation that the theory of evolution by natural selection is supposed to provide: a fact about a change in the distribution of traits in a population. If we deny (2) on the grounds that this explanation is somehow “superseded” by an alternative explanation appealing only to the traits of Paul’s parents and the laws of heredity, it seems that we would also have to reject this latter explanation on the grounds that it would be similarly superseded by an explanation appealing only to the microphysical details surrounding Paul’s conception and the laws of microphysics.

is, it must be a process that, *ceteris paribus*, produces configurations and avoids others in a way that is responsive to the fact that the ones produced are better for achieving the intended effect.

Learning, too, seems to fit the comparative conduciveness explanatory pattern. Suppose Boris the black bear learns how to open a trash can whose design just barely undershoots the famous narrow range between the abilities of the most capable bear and those of the least capable human. In any case we would be inclined to count as genuine learning, Boris's behavior will, as a result of his experience, have come to exhibit certain features rather than others because these former better conduce to trashcan-opening.

5.2 – Ruling out Boorse's Knockout

In Boorse's Knockout, no comparison between propensities is relevant. The leaky setup sticks around because it *does* keep the scientist unconscious but it's not relevant that it is *better* at keeping the scientist unconscious than some alternative.

To see this, suppose that the closed setup had possessed, for some reason, an even higher propensity to make scientists unconscious than the leaky setup does. The scientist would probably have become unconscious earlier due to the closed setup. But then, when the leak formed, the scientist would remain unconscious and the leak would persist. It wouldn't spontaneously repair itself just because the closed setup was even more unconscious-making. The setup isn't subject to any process "attracting" it, *ceteris paribus*, to whatever configuration is more scientist-unconsciousness-conducting. The direction of the "propensity-gradient" between the closed and leaky setups has flipped, but the result remains the same.

5.3 – A more Difficult Test for CCET

Here is a tougher case for CCET.

Strong Boorse: Add two additional details to those of Boorse’s Knockout. First, suppose that the experiment the scientist is conducting requires her to operate a manual pump that causes strain on the hose. Second, suppose that the strain caused by her pumping causes the initial formation of the hole.

Intuitively, this should still be a non-teleological case. But the counterfactual story I just told to verify that CCET does not apply to Boorse’s Knockout no longer works. In Strong Boorse, if the closed setup had somehow had high propensity to make scientists unconscious, higher than that of the leaky setup, this *would* change the outcome of the case. The closed setup would have knocked the scientist out before she could cause the leak, so the leak would never have formed.

In Strong Boorse, then, it matters both that the closed setup has a low or no propensity to keep scientists unconscious and that the leaky setup has a significant propensity to do so. But this does not mean that CCET ascribes a function. A fact about the specific pair of propensity values of the closed and leaky setups, namely

Value-Pair (VP): The leaky setup has a substantial propensity to make scientists unconscious and the closed setup has little to no propensity to do so

is relevant to why the setup remains leaky. But it does not follow that the fact of the *direction of the inequality* relating these propensity values, namely

Inequality-Direction (ID): The leaky setup has a greater propensity to make (is more conducive to making) the scientist unconscious than the closed setup

is explanatorily relevant to the setup’s remaining leaky.

It might sound *ad hoc* or suspicious, at first, to suppose that the leak could persist because of VP, without its also persisting because of ID. After all, isn’t ID an immediate consequence of VP? But with a moment of the reader’s indulgence, I can show this is the right verdict.

Inequality-Direction's merely being a consequence of Value-Pair doesn't automatically mean that it explains whatever Value-Pair explains. If whenever B explained A, every consequence of B also explained A, then the fact that 7 is greater than 6 would explain everything.

Further, there are situations in which it is clearly a specific pair of values, and not the direction of the inequality relating these values, that is explanatorily relevant. For example:

Waitlist: At a full restaurant, the Hatfields, a party of 2, arrive first, and the McCoys, a party of 4, arrive later. Nonetheless, the first table to become available is a 4-top, and the McCoys are seated first.

Here, the McCoys eat first because of the specific pair of numbers 2 and 4, and not, as such, because of being the larger party. Had the Hatfields brought a party of 3 instead of 4, the inequality's direction would have remained the same, but the Hatfields would have been seated first since a party of 3 gets a 4-top.

Analogously, it is Value-Pair that explains why the leak remains in Strong Boorse, and not Inequality-Direction. To see this, change the propensity values without changing the direction of the inequality and watch the outcome of the case change. For example, suppose that right when the scientist began the experiment, God changed the principles of human physiology so that nitrogen had a high propensity to make humans unconscious, though a mixture of nitrogen and chlorine had an even higher one. Then the scientist gets knocked unconscious before she can cause the leak. The leaky setup still has a greater propensity to make scientists unconscious, yet the setup never develops a leak. What the process that formed the leak "cared about" was not a gradient in propensity-values as such, but specifically that the closed setup doesn't make scientists unconscious and the leaky setup does.

§6 – Teleology Made Easy – Appearance-as-Guide

CCET is a distinctively liberal etiological theory of functions. By distinguishing between functionary and beneficiary and requiring only that the beneficiary and not the functionary bear the mark of purposive organization, it allows more routes to a function attribution than etiological theories without this distinction do. Moreover, the etiological condition requires the mere existence of *a* mark of purposive organization, allowing many, often redundant paths to a function attribution, as evolved and designed things often possess many marks of purposive organization. The result, as I will discuss, is that a spandrel *s* is likely to have its apparent function despite its spandrelhood. Though *s* itself, by definition, not an adaptation for its apparent function, it possesses its apparent functions as an *organic* function, the relevant marks of purposive organization being (perhaps very minor) *complementary* adaptations elsewhere in the organism which promote the functioning of *s*.

Spandrels lack *proper* functions in my sense of the term. But typically, when we are asking about the function of an organism's trait, we are interested in its *organic* function, what it is for relative to the *organism*. Thus, it need not be the evolutionary history of the trait itself that grounds its purposiveness. Because it is the beneficiary – the entire organism – that must bear the mark of purposive organization, the function of a trait with respect to an organism may derive from the etiology of complementary traits.

Consider Gould and Vrba's example (1981), the black heron. Black herons use their wings to mantle: to huddle over the water they stand in, casting a shadow on the fish below them. There's a clear sense in which mantling seems to be one of the things the black heron's wings are for; it's one of the uses to which she characteristically puts them. But they did not arise by being selected for mantling. Nor are they retained due to the contribution of selective pressure to mantle. Nor are they likely to have been optimized for mantling since shadow-casting is presumably orders of magnitude

less finicky about wing shape than is flying. The wings are therefore spandrels, in this chapter's sense, with respect to the function of mantling.

To see how the black heron is organized around the purpose of using its wings to mantle, look to the evolutionary history of its *other* traits, and in particular, its brain. In some population of ancestors of today's black herons, there were birds with bird-brains which were, under the right circumstances, apt to cause something roughly like mantling behavior with the wings, and there were birds whose cognitive setups were not so apt to have this effect. Over time, the cognitive basis of this aptitude came to predominate because of its greater tendency than alternative cognitive setups to conduce to the useful behavior of mantling.⁹ This trait of the population in turn explains why the brain of a black heron born today will develop the cognitive basis of mantling. The direction of a bit of ontogeny today is explained by the historical pressure for such a bit of ontogeny to go in a way that will have a particular later effect: mantling.¹⁰ Thus, the use of the wings to mantle figures as a final cause in explaining something about how today's baby black herons' lives unfold.

Allowing the attribution of functions on the basis of complementary trait selection reduces the risks that spandrel-hood poses to ascriptions of purposiveness to near negligibility. Consider the worst-case scenario about the prevalence of spandrels in nature. Jerry Fodor and Massimo Piatelli-Palmarini's *What Darwin Got Wrong* (2011) argues that natural selection plays only a marginal role in explaining the course of evolutionary history. According to Fodor and Piatelli-Palmarini, the great majority of macro-level traits arise and become fixed by constraints besides selection for performing

⁹ Justin Garson helpfully points out that, in all likelihood, ontogenetic learning, and not just instinct, plays a role in the reliable effective production of mantling behavior. All the same, black herons mantle and other birds who perfectly well could do not; it is not a *tabula rasa* phenomenon. Once some learning to mantle has gone on, this provides an additional route to the wings being for mantling. But it is in virtue of such traits as the basis of the ability to (learn to) mantle that the wings have this function not only after some life experience, but from birth.

¹⁰ Suppose that even the rudimentary basis of the mantling instinct is a spandrel. Never fear: just make the same argument with respect to the basis of this instinct's robustness or fine-tuning.

this or that function. Natural selection is left, on their view, to play a minimal role at the end of such a process, one of “tuning the keys.” Fodor and Piatelli-Palmarini’s position is controversial, even unpopular among biologists and philosophers of biology. What is important here is that even if it is correct, it poses little threat to the idea that biological traits are generally for what they are characteristically used for. As long as selection for o’s doing π explains even minute fine-tunings or stabilizations for complementarity of other parts of the organism, o’s doing π is acting as a final cause.

I’ll drive the point home with some imagery. Compare the flow of matter through an individual organism’s lifespan to the flow of a river. The characteristic changes that an organism of its kind undergoes as it develops and ages, and in its day-to-day gyrations, make up a winding riverbed path which this flow of matter follows. Fodor and Piatelli-Palmarini’s pessimistic view about the importance of natural selection suggests that these riverbed paths are not generally initiated by natural selection. The flow of life for a species is channeled into such a pattern by other constraints. Nonetheless, there is a role for natural selection as a fine tuner, to dig a species in to ariverbed, and entrench the pattern of flow exhibited by its life. It is this digging-in and entrenching – the fine-tuning creatures to do better what they already do – in virtue of which, even in this minimal role, natural selection serves as a genuine final cause, shaping biological entities to accomplish those aims which are manifest in their organization.

Chapter 3 – A Guided Tour of the Teleological Landscape

The goal of this chapter is to provide an account of what it is to have a belief, and what it is to be subject to a requirement in terms of which it can then be argued that our beliefs are subject to the requirements of truth-normativity. My approach begins from the framework of the theory of purposes I advanced in the previous chapter. My attempt to bake deontic and doxastic cakes out teleological ingredients requires a number of intermediate steps. By way of introduction, I will give some characterizations of the general strategy behind my approach, and then a roadmap of how it proceeds.

The basic premise of my attempt here is that there is a level of organization at which a great many of the creatures that our common sense thought deals with in navigating our lives live: the level of commonsense teleology. Questions about what things are for, of what would further their purposes, of what means they have at their disposal to accomplish these purposes, are relevant to a great many and a wide variety of our dealings with the world. The conceit of the effort undertaken here is that the phenomena of belief and requirement can be found at this level. Importantly, the level of organization with which commonsense teleological thought deals is not simply a matter of an agglomeration of purposes. We thus cannot describe it if our only theoretical notion is that of a purpose – we can't substantively characterize the teleological landscape in the ways we need to simply by endlessly listing purposes. We need a battery of related notions.

After some preliminary development of expressive tools useful for expressing the constituent notions in section one, I will proceed as follows. First, I will develop a notion of *normal performance*, which is to act as a kind of “glue” unifying the performance of a functionary that does π when it is meant to do π , when this is no accident. Deploying the notion of normal performance, I will define a pair of notions that describe the teleological situation of any functionary subject to purposive organization. These are its *remit*, the set of possible demands that can arise out of the

purposes governing it in the course of normal functioning; and its *repertoire*, the set of characteristic activities at its disposal by which it may meet the demands of its remit.

In order to develop a notion of requirement and of belief, we need a concept of a way that something's situation can "point" at some circumstance (actual or non-actual) in a way that is more exclusionary – less permissive – than merely having an aim or a purpose. To fill this gap, I introduce a notion of some activity's being an *effort* by some functionary directed at some further activity.

In terms of remit, repertoire, the notion of effort, and the notion of an indicative intentional signal (which I introduce as undefined, and define later on), I propose a way of understanding the relation between activities expressed in a hypothetical imperative, a notion of some activity's being *called-for* by some other. My general approach to this task involves taking the idea of the predicament some functionary faces in instructing another functionary in the performance of some task and abstracting away from the specific preferences or other idiosyncrasies of the giver of commands, to think about what features of the situation and task itself make certain commands fitted. Here the idea will be that an umbrella activity Π calls for some activity ϕ as a means when there is some point in doing ϕ relative to furtherance of the aims involved in doing Π , that there is some way of doing ϕ without compromising one's other aims, and there is some way of inducing an effort to do ϕ by means of a signal instructing the functionary bound by the conditional requirement to do ϕ .

I go on to advance an account of intentional signals, again using the basic ingredients that describe a functionary's predicament, its remit and its repertoire, along with the notion of an effort. I then go on to define a belief that P as a condition of preperadness to signal P in the relevant context, where to be prepared is the result of a certain sort of effort, a kind of "teleologically enriched disposition."

Before diving in, I want to make one point of “pre-tour” orientation. I will be talking a lot, in this chapter and the next, about what representation and communication do. The claims I will make should always be interpreted along the lines of what Dennett (1989) following Chomsky (1980) calls a “competence model” as opposed to a “performance model.” My claims will be about some of the things – among many others that are not my focus here – that our cognition and speech do for those who can engage in them. They are not intended to make any commitment about the fine or basic structure of how they manage to do these things for us. For instance, natural language is clearly compositional, and thought is generally taken to be as well – but there will be hardly a mention here of proposition-expressing thoughts or bits of language being made up of components. My main concern is to propose what it is we accomplish in thinking and saying, regardless of how we manage to accomplish it.

§1 – Plans and Planning Connectives

In the previous chapter, I defended the following account of purposes:

Comparative Conduciveness Etiological Theory (CCET): *With respect to X, o is for doing π just in case, for some properties F and F*, X is F rather than F* because X’s being F is more conducive to such things as o doing π in some conditions C than would be its being F*.*

I begin my discussion here with a way of answering the question: what sort of thing is ‘ π ’? There are two ways of reading this question. There is the question: “these things which are called ‘ π ’ in the definition of CCET: what are they?” This is a sort of ontological question about purposes.

“This expression ‘ π ’ which refers to the purpose in a purpose attribution: what is it?” This is a sort of meta-linguistic question about the expressions that pick purposes out. I address the ontological question by way of the meta-linguistic expression. Activity-kinds that can be purposes,

or *plans* as I will call them, are those things designated by the members of the system of expressions I will describe below.

What I will propose, effectively is a way of generating expressions that refer to arbitrarily complex activities, which involve or relate to simpler activities, one which perspicuously indicates how the complex activity in question involves or relates to those simpler activities.

To begin to answer the question “what sorts of expressions can go in the place of ‘ π -ing?” we can begin with gerund expressions of natural language. The sorts of things that something may be for include, for example, “flying around,” “digesting fat,” or “seeing through walls.”

There are, however, many purpose-attributions we would like to make but which would be awkwardly expressed using natural language gerunds. To see why, notice that purposes tend to come in groups. Complex multicellular organisms have many body systems meant to do a wide range of different things. And these various purposes do not inhere together in an organism as an unstructured pile; they are interrelated in various specific ways. The point iterates if we think about body systems themselves, and the various organs which make up those systems, to say nothing of the cells which make up the tissues of those organs and the many components and products of each individual cell. There are various parts meant to do interrelated things. By the same token, the whole system is meant to function in a way that involves its various parts doing various things in specifically interrelated ways. Much the same point holds of the various aspects of the socio-cultural-technological apparatuses which populate our life-world.

There are two points here about the kind of complexity of purposive organization we want to be able to articulate in purpose attributions: a variety of simpler purposes arranged in a structured way, and a variety of parts of a functionary to which these component purposes belong. Consider for example, the question of what the endocrine system is meant to do, for the purposes of the

organism of which it is a part. Here we have organism as beneficiary and endocrine system as functionary. If we want to say with any specificity what the endocrine system is for doing, we have to say quite a lot, and to compress all of it into a single gerund expression would be unwieldy at best. As a functionary, the endocrine system is comprised of many parts of diverse kinds – for example the thyroid, and adrenaline – and different sorts of activity belong to different sorts of parts in interrelated ways. So we would like to be able to articulate purposes in a way that assigns different component activities to different component parts of the functionary, as well as indicates the way in which the purpose in question means for these to relate.

Below I propose an expressive toolkit for building expressions for complex structures of purposes, which I call “plans,” which govern multipartite functionaries, out of component expressions which designate simpler component purposes and assign them to component parts of the functionary.

These expressive devices work in a way that is analogous to how logical connectives build complex sentential expressions out of simpler ones. Because of the analogy in how they work, I call them “planning connectives.”

Simple Achievables: A simple achievable is the activity denoted by an expression ‘ π'_a ’ which is not built out of constituent expressions by the application of planning connectives as described below. *Swimming, running, seeing through walls*, and the like are simple achievables. ‘Swimming,’ ‘running’ and ‘seeing through walls’ are expressions designating simple achievables.

We can read $\pi(o)$ as “o does π ”. If we want to attribute the purpose doing π to o , relative to some X , we can say that the plan π is *operative* for o relative to X . We can write $OP^X\langle\pi(o)\rangle$ for short.

I now present five connectives, by which we can build designations of complex activities of composite things out of designations of simpler activities of simpler things. These complex activities will be specified by specifying their eventuation conditions. Their eventuation conditions will be specified as a function, corresponding to the connective, which takes the eventuation conditions of the component plans as inputs.

1.1 – Planning Connectives

The first three are conjunctionlike:

Simple Conjunction: ‘&’

Let π_a and π_b be plan expressions and let u and v denote individuals and let ' O ' denote an individual of which the mereological sum $u \oplus v$ is a part.

The plan $\pi_a(u) \& \pi_b(v)$ transpires for or o just in case π_a transpires for u and π_b transpires for v .

For simpler notation, when u , v , and o are identical, we can write simply $[\pi_a \& \pi_b](o)$

Example: Suppose, for example, that a crane, *Frasier*, flaps his wings, (call the wings $w_{Frasier}$) and shakes his tailfeathers, ($t_{Frasier}$). If we want to express what transpired using the planning formalism, letting π_{fw} designate *flapping one's wings* and letting π_{st} designate *shaking one's tailfeathers* we can write $[\pi_{fw} \& \pi_{st}](Frasier)$

We can deploy these expressions within the formulation of the CCET to formulate function ascriptions. When a plan expression describes what a functionary is supposed to do (relative to some X , Suppose a scientist surgically implants a device in Frazier's brain, with the idea that the device will get him to flap his wings and shake his tailfeathers. If we want to say that $\pi =$ *flapping(w)&shaking(t)*, the flapping of the wings and the shaking of the tailfeathers, is what Frazier is meant to do, relative to the purposes of the surgery, we can say

$OP^{SURGERY} \langle \pi(\text{Frazier}) \rangle,$

Or, spelling out π in terms of its compositional structure:

$OP^{SURGERY} \langle \pi_{fw} \& \pi_{st}(\text{Frasier}) \rangle,$

To indicate that the surgery has certain features rather than others because the former features are more conducive to $\pi(\text{Frazier})$, i.e. $flapping\&shaking \langle w, t \rangle(\text{Frazier})$, i.e. Frazier's flapping his wings and shaking his tailfeathers.

1.2 – Means-end Conjunction

Suppose that the surgery performed on Frazier was part of a captivity breeding program, conceived under the theory that his dancing will get the attention the zoo's female crane, Diane. It is an aim of the program that Frasier flap his wings and shake his tailfeathers:

$[flapping(w_{Frasier})\&shaking(t_{Frasier})](Frasier)$. But it is also an aim that Diane become interested in mating with Frasier: $int_F(Diane)$. But these events are meant to transpire for the prospective couple in a particular way, the latter is supposed to happen *because of* the former happening. We can denote this relation between the elements of a plan with \gg , or *means-ends conjunction*.

That is, if Diane becomes interested in mating with Frasier as a result of his dance, we can write a plan expression of the form $(\pi_a \& \pi_b) \gg \pi_c$. In this case, we can write:

$[flapping\&shaking(Fras) \gg int_F(Dian.)]$

to describe the fact that these are the events which are meant to transpire, relative to the purposes of the program, we may write:

$$OP^{PROGRAM}(\pi_{fw} \& \pi_{st}(Fras.) \gg \pi_{iF}(Diane))$$

This is true just in case the program takes a certain form rather than another because this form better conduces to things transpiring with the couple in the relevant way: that is, that Frasier flaps his wings and shakes his tail feathers, and by his so doing, the exciting the passions of Diane.

Note that the connective \gg is general as between causal and constitutive means, so both the plan of breaking a window by throwing a rock and a plan of scoring a run by hitting a ball over the fence in fair territory are plans of the form $\pi_a \gg \pi_b$.

1.3 – Succession

Suppose that the captivity breeding program is designed to involve a study on the effects of Frasier’s dance on subsequent mating behavior, and it is part of the experimental design that Diane and Frasier should not have mated prior to Frasier’s performing his dance. Then, it is among the aims of the program that $\pi_{fw} \& \pi_{st}(Fras.) \gg \pi_{iF}(Dian.)$ transpired, and it is among the aims of the program that mating transpire for the couple, that $[\pi_m(Couple)]$. But these aims are not held in a way that is unstructured with respect to the way they are meant to transpire in relation to one another. Rather, it’s the aim of the programs that they transpire in a specific order in time. We can represent this structure - conjunction with the added constraint of succession – with the connective: ‘ \succ ’

So we can write $OP^{PROGRAM}[\pi_{fw} \& \pi_{st}(Fras.) \succ \pi_m(Coup.)]$

Generally,

Definition (>): For any plans π_a and π_b say that $[\pi_a(O_1) > \pi_b(O_2)]$ iff

- 1) at some time t_1 , $[\pi_a(O_1)]$ while $\sim[\pi_b(O_2)]$ (π_a has transpired for O_1 while π_b has not transpired for O_2) and
- 2) at some later time t_2 , both $\pi_a(O_1)$ and $\pi_b(O_2)$ (π_a has transpired for O_1 and π_b has transpired for O_2).

The expression:

$$OP^X[\pi_a(O_1) > \pi_b(O_2)]$$

Is thus to be read as saying that for X 's purposes O_1 and O_2 are meant to function as follows: O_1 does π_a and then O_2 does π_b .

Observe that the two kinds of special conjunction, means-end and succession, may overlap: we might plan for Abel to leave a note reading “Bertha: press the yellow button!” and for Bertha to then press the yellow button, both after and because of Abel’s note.

Note also that we may have one without the other: a bird who tokens the neural basis of the song that goes *tweedly-dee* may be meant to sing *tweedly* and then to sing *dee*, but not particularly to do the latter *because* he did the former.

Conversely, we can have means-end conjunction without succession. A plan may call for destroying some documents by shredding them, but it does not thereby call for shredding the documents *and then* destroying them.

We can define a “hybrid” connective to specify the first sort of case, in statements about what transpires, that succession and means-end conjunction occur together, and in statements about governing purposes or plans, that they are meant to occur together.

1.4 – Contingency Plans

Consider the mimic octopus, which uses color-changing cells called chromatophores to adopt the appearance of both the ocean floor on which it sits, for camouflage, as well as the appearance of the other animals in its environment for various reasons: it has been reported to mimic the conspecifics of prey animals seemingly in order to predate them, to mimic predators of predators, in order to deter them, as well as to mimic predators of animals who pose nuisances in order to dispense with their nuisance-making.

Suppose that such an octopus has learned to engage in such a behavior in a situationally sensitive way. We'd like to describe the manner of functioning around which it is organized as a complex plan involving multiple constituent plans as parts: namely, mimicking the conspecific of an animal and mimicking the predator of an animal. But the octopus is not organized around the aim of doing these things in a way that is unstructured with respect to when and how.

Rather, we would like to understand the behavior of the octopus as a contingency plan, in which the functions described in the various constituent plans are meant to occur in different situations: when a nuisance is present, mimic a predator of that nuisance, and when prey is present, mimic that prey's conspecific, and when a predator is present mimic a predator of that predator.

We are used to expressions for logical connectives as tools for building expressions out of two simpler expressions. To express kinds of structured patterns of activity such as that around which the mimic octopus from above is organized, I introduce a notational tool for construction of expressions of plans out of expressions for *functions* mapping propositions to plans. Such functions may have domains with any number of propositions and may range over any number of plans. The basis for this tool will be the expression ' \mapsto ' (pronounced "maps to") which is one of the notational devices commonly used to specify functions in mathematics. When a function operates on a small

domain of inputs, we can use ‘ \mapsto ’ to specify the exact identity of the function, specifying each of its inputs and each of its outputs. For instance, the expression for the function which has the members of my committee as inputs and the floor of their Emerson Hall offices as outputs may fully specified as follows: $\{RM \mapsto 3, SS \mapsto 3, MR \mapsto 3\}$.

When a function is defined on a larger but still finite range of inputs, we can use ‘ \mapsto ’ to specify it elliptically. So for example the English ordinal scheme for assigning numbers to letters of the alphabet in which a maps to 1, b maps to 2, and so on, can be expressed as follows:

$$\{a \mapsto 1, b \mapsto 2, \dots y \mapsto 25, z \mapsto 26\}.$$

When a function is defined on an infinite domain of inputs, we may be able to express it schematically using ‘ \mapsto ’. For instance, we may express the function which maps from a natural number n to the number resulting from multiplying n by 2 as follows:

$$\{n \mapsto 2n\}$$

We can then express the aim around which the mimic octopus is organized using an expression designating a function from situations to the activities planned for that situation:

$$\{Prey\ animal\ present \mapsto \pi_{consp.}, Predator\ present \mapsto \pi_{pred.}, Nuisance\ present \mapsto \pi_{pred.}\}(Paul),$$

Correspondingly we can represent the fact that this plan is operative for Paul the octopus (as a proper function, i.e. relative to his own purposes) with the expression

$$OP^{Paul}\{\{Prey\ animal\ present \mapsto \pi_{consp.}, Predator\ present \mapsto \pi_{pred.}, Nuisance\ present \mapsto \pi_{pred.}\}(Paul)\}$$

The kind of organization naturally expressed as a contingency plan is not limited to the basis of what we'd typically call intelligent behavior. Consider the action of the bicarbonate buffer system in your blood. It is meant to function by neutralizing acids when they are introduced to the blood, as well as by neutralizing bases when they are introduced.

More generally we can define contingency-plan expressions as follows:

Definition (\mapsto): For any propositions $P_1, P_2 \dots$ and plans π_1, π_2, \dots , and individuals O_1, O_2, \dots , and U of which each O_i is a part, say that $\kappa(U) = [P_1 \mapsto \pi_1(o_1), P_2 \mapsto \pi_2(o_2), \dots](U)$ ¹ just in case one of the P_i is true, and for each i either $\sim P_i$ or $\pi_i(O_i)$.

Again, by applying the CCET's notion of purpose, we can use 'OP' to describe operative contingency plans. So for example, we can read $OP^X\langle [P_1 \mapsto \pi_1(o_1), P_2 \mapsto \pi_2(o_2), \dots](U) \rangle$ as ascribing, relative to the purposes of beneficiary X, governance of the functionary U by the contingency plan that U 's constituent o_1 do π_1 if P_1 is true, that U 's constituent o_2 do π_2 if P_2 is true, and so on.

Call each P_i in a contingency plan κ , a *contingency* of κ and call the corresponding π_i the *branch* of κ corresponding to P_i . When someone functionary caused to pursue the branch of a contingency plan corresponding to that contingency, say that it is *adapted* to that contingency.

In some cases, when a contingency plan is operative and becomes adapted to a specific contingency, the corresponding branch of the plan becomes operative as an aim. So for example, if you intend to order for someone arriving late to a restaurant, there is, in virtue of your intention, operative for you the contingency plan of ordering a dish in the circumstance that your friend requests it. If your friend tells you they want the Peking duck, then you form the intention to order

¹ (in the special case where there is only one individual involved, i.e. $O_1 = O_2 = \dots = U$, we can write $[P_1 \mapsto \pi_1, P_2 \mapsto \pi_2, \dots](U)$).

the Peking duck, and the plan to do so – a branch of the original contingency plan – is now operative for you.

In other cases, however, a functionary may be driven down a branch of a contingency plan without that branch becoming operative. When acid is introduced into your bloodstream, the bicarbonate buffer system in your blood neutralizes it. But this need not involve any specific purposive organization specifically related to the neutralization of acid, over and beyond the general purposive organization around the contingency plan of neutralizing both acids and bases.

§2 – On Being Normal

Function attributions draw a kind of distinction between the good case and the bad case in a very straightforward way: the good case is when σ does π and the bad case is when it does not. But there are further distinctions between cases which are good in just this minimal sense, on the one hand, and cases which are good in some stronger sense on the other. These distinctions between the good version of the good cases and the not-so-good versions of good cases are generally connected with the idea of a distinction between an accidental and a non-accidental kind of correspondence between what purposes dictate lining up with what actually happens in an accidental way. When the fit is non-accidental, things are *normal*, when it is accidental, things are *abnormal*.

Following Millikan in *LTOBC*, Millikan's we can make two sorts of important distinctions between the normal and the abnormal. First is the distinction between normal and abnormal *conditions* for the performance of a function, and between normal and abnormal *performance* of a function – sometimes also referred to as performance of the function in accordance with a normal explanation. These notions, and especially the latter, will be of central importance my project as well.

The basic idea of some functionary's being normal *conditions*, as I would like us to think of it, is the idea of conditions matching those that something evolved or was designed (or both) to function in.² Falling outside of these conditions can attenuate the sense in which the "success" of something's acting in a way that coincides with its function would really be a success. Sprinkler systems are built to douse the buildings they are installed in with water; but they're not built to do this on perfectly uneventful days. An umbrella is built to keep things directly beneath it dry; but it's not built to do this on the ocean floor.

The basic idea of some functionary's *performing* its function normally is the idea of something's performing its function in the *way* that it evolved or was designed to perform its function. If your alarm clock falls on your head and wakes you up, it performs its function abnormally: it does what it is designed to do, but not in the way that it is designed to do it.

Though I am about to criticize Millikan's approach to normalcy, I nonetheless adopt from her an emphasis on the importance of the notion in relating the biological to the intentional. Any tract of biological activity, like any tract of any activity, can be described as conforming to a virtually unlimited variety of distinct patterns. Any scheme worth having for describing a subset of biological activity in *intentional* terms must be able to distinguish between those descriptions under which the activity was intentional and those descriptions with which the activity merely coincided. Further, I want to pick out such descriptions in a way that doesn't go in a circle and beg the question with respect to the nature of intentional relations, we need some broader category of intentional, but not yet intentional relations between tracts of activity and descriptions under which they fall out of which to construct intentional distinctions.

² This gloss doesn't respect the distinction between functionary and beneficiary, but the way of adapting it so that it does is unwieldy: as a functionary, being in normal conditions to do π is being in conditions that whatever thing evolved or was designed so that you do π evolved or was designed for you to function in.

The notion of an operative plan is a starting point; but it is not on its own a sufficient basis for reconstructing intentional characterizations of individual tracts of activity. There is doing π in keeping with the purpose that you do π ; and then there is merely doing π *while* it happens to be something's purpose that you do π , but only coincidentally rather than in keeping with this purpose. which satisfies the purpose because of and in keeping with its organization around this purpose. What we are looking for is a kind of unity connecting a particular tract of activity and a description under which it falls which is akin to the relation a tract of *rule-following* activity has to the rule it follows, but one which we can articulate in non-intentional (or more specifically, intensional-but-not-intentional) terms. The notion of some actor's performing its function normally, if it can be spelled out in non-intentional terms, fits this role of "rule-following lite," and it is this role which I think has made the notion of normalcy a rightful object of central concern in biological approaches to intentionality.

In the remainder of this section, I will criticize what I see as the two extant proposals for something like the notion of normal or non-accidental performance of a teleo-function: Ernest Sosa's conception of *apt* performance, and Millikan's notion of Normal performance. I will then propose my own notion.

2.1 – Against Sosa's Approach

In several works (2007, 2011) Ernest Sosa has advanced a virtue-epistemological framework that emphasizes that beliefs are performances, and that performances are subject to the 'AAA' framework of evaluation. 'AAA' is an acronym for three respects in which a performance may be evaluated either positively or negatively: with respect to its accuracy – whether it in fact does what it aims at; with respect to its adroitness – whether the performance manifests a general capability to produce accurate performances, irrespective of whether this performance itself happens to be

accurate; and whether it is apt, whether the performance was accurate, and adroit, and its accuracy and adroitness – its success and its skillfulness – stand in the right sort of non-accidental relation.

Sosa's notion of apt performance, then is a candidate to play the role of what I am calling normal performance of a teleo-function. Sometimes Sosa describes a performance's aptness as a matter of its accuracy being a *manifestation* of its adroitness. This seems to capture the notion, but it raises the question under what circumstances we should understand the accuracy of a performance as manifesting, rather than merely coinciding with, its adroitness. One may worry that this is a redescription, rather than an explanation of the notion.³

At other times, Sosa describes the aptness of a performance as a matter of its being accurate *because* it is adroit. This second suggestion offers what seems more like an attempt at analyzing the notion of aptness. But I think that it cannot be correct as a definition of the kind of non-accidentality of success that we want to capture. The problem is that this definition allows for deviant causal chains. Suppose an archer takes aim at a target – a favorite example of Sosa's – and suppose that unbeknownst to her, the archery range is the site of an experiment testing a network of nano-bots designed to manipulate wind-currents. Suppose that these nanobots are operated by an expert archery instructor, who can tell good technique from bad technique – adroit shots from maladroit ones – and who drives the nanobots to blow arrows that are well-shot to the target – though not *via* a straight path – and arrows that are poorly shot away from the target. Then the archer's shot will be accurate and adroit, and it will be accurate because it is adroit. But its success will nonetheless be accidentally related to its adroitness *qua* adroitness. Rather, its success is

³ In fact, it may be that all that Sosa intends to do is to describe the notion in a way that fixes the readers mind on it. The discussion of how the notion of aptness is to be understood passes by very quickly when he introduces it, and his main concern does not seem to be to defend the reality and coherence of the notion, but to begin with the hypothesis that it is in good standing and make use of it in epistemology.

mediated by and dependent upon the fact – accidental relative to the archer’s capability – that the instructor driving the nanobots elects to reward adroit shots and punish maladroit ones.

2.2 – Against Millikan’s Approach

Millikan’s approach to the notion of normalcy centers around the idea of the normal *explanation* for the performance of a proper function. Let R be a group of individuals (including bygone individuals) sharing a function F in virtue of their genealogical relation. For Millikan, the *most proximate* normal explanation of how members of R perform F is (1984, 33)

the *least detailed* explanation possible that starts by noting some features of the structure of members of R, adds some conditions in which R has historically been when it actually performed F – these conditions being uniform over as large a number of historical cases as possible – adds natural laws, and deduces, i.e., shows in detail without gaps, how this setup leads to the performance of F.

The (most proximate) normal conditions are those conditions cited in such an explanation. (Most proximately) normal performance is performance in keeping with the most proximate normal explanation. Less proximate normal explanations trace things further back, to the historically preponderant causes of the normal conditions cited in the most proximate normal explanation. So, for example, the most proximate normal explanation of the pumping action of a heart will probably cite the input of an electrical signal, and a less proximate one will specify that this signal normally comes from the brain.

I doubt that this approach adequately identifies a normal-abnormal distinction. The main problem for this approach is its requirement that there be an explanation that is historically preponderant while also deductively entailing that the entities in question perform their functions. That this should be a problem is especially clear in cases of functional kinds whose members rarely perform their functions. It is a function of an animal sperm cell to fertilize an ovum. Only a small

fraction of sperm cells actually manage to perform this function. If we explain in general terms how those that do fertilize an ovum manage to do so, we will cite conditions (e.g. proximity to an ovum) and properties (e.g. the structural basis for its motility) that successful sperm cells share with many counterparts that do not perform their function. So such an explanation does not show deductively how such a function gets performed. If we descend down to citing the particular details of how the successful sperm cells have succeeded where their counterparts failed – of how exactly, in each case such-and-such a particular sperm cell got where it was going first, and its counterparts were too late – we will get an “explanation” that is wildly disjunctive, with the precise details differing in every case, and one that is not projectible for future cases at that. There is no non-circularly identifiable set of properties that a preponderance of successful sperm cells share and all unsuccessful sperm cells lack. Perfectly normal performance in perfectly normal circumstances does not guarantee success.

2.3 – My Proposal

The notion of normal performance takes a little more work to hammer out. The basic idea is this. Whenever some X is somehow adapted to conduce to σ 's doing π , there is some *way* of σ 's doing π to which X is adapted. Performing normally is just performing in this way: that way of σ 's doing π to which X has been adapted. More specifically, it is performing in that way of σ 's doing π to which X 's being F conduces in virtue of which X has F -ness as a mark of purposive organization around σ 's doing π .

Something may perform abnormally because it does π in a way to which its marks of purposive organization simply do not conduce. For example, the wiring in your alarm clock is there because it helps the clock wake you up on time, but the wiring is irrelevant, not conducive, to its waking you up by falling on your head.

But a functionary may also perform abnormally even when it performs its function in a way that its marks of purposive organization *do* conduce to. It does so when it performs its function π in a way that, though its marks of purposive organization conduce to performing π in this way, is nonetheless unrelated to how it got them. Suppose an antibody your body produced in response to an infection figures in causing the infection to subside, but it does so not by interacting with the pathogen in your body but by interacting with a reagent in a test tube, causing you to be diagnosed and subsequently cured. The antibody's structure is a mark of purposive organization, adapted to the specific pathogen that has infected you; and it is just this feature that causes the correct diagnosis and your being cured. But the fact that your antibody was apt to help cure you in this way was not *why* it got that structure.

The preliminary characterization identifies normal performance with *the way* of performing the function to which the beneficiaries' marks of purposive organization conduce in virtue of which it has said marks. This won't do, however, because of the ambiguity in what counts as *the way* that a beneficiary has evolved or learned or been designed to work. With products of natural selection, for instance, the ambiguity arises as follows. We are faced with a particular successful performance and ask: is it of-a-kind with the successful performances of the functionary's forebears? Here we are asked to look at a historical set of particular successes, to make a generalization about the way in which these successes occurred, and to check whether the new success fits within our generalization. But our first-pass definition provides no guidance about *how* to generalize from the particular historical precedents. It therefore allows "cheating" in how we generalize about past successes.

Cheating might involve describing the evolutionarily significant past successes overly narrowly, to guarantee the verdict that a new success is abnormal, even if not warranted: someone might say that the lungs of a mountain-climber using supplemental oxygen work abnormally because

their ancestors' lungs evolved to oxygenate blood by breathing *air*, not by breathing supplemental oxygen. But this verdict runs counter to the aim of defining normal performance to capture a notion of non-accidental success. The intuition I want to evoke here – and want my notion of normalcy to fit with – is that the lungs may need extra help at the top of Everest, but in the good case, they're still doing their part.

Cheating might also involve generalizing too broadly from the set of past precedents to guarantee the verdict that a given performance is normal, even when we want our notion of normalcy to give the opposite verdict. Here is a fanciful but clean case. Suppose an owl flaps its wings, and that it causes, butterfly-effect style, a cooling wind to blow through its neck of the woods on a hot day a month later. The owl's feathers are for regulating its temperature. And they do just that in this case. But, we want to say, they do this not in the way they evolved to, but accidentally. But what stops us from cheating by describing the normal way that the owls feathers accomplish this function exquisitely loosely? What stops us from saying that in the owl's ancestors, the feathers regulated the temperature by making a difference to the flow of air around the body, and in this case, too, they did just that?

Cheating in both directions may also be possible in describing the way in which intelligently designed artifacts are meant to work. Shine a flashlight out of the porthole of a submarine and spot a fish; a would-be cheater might try to claim that the flashlight performs abnormally in making the fish visible to you because the designer only had in mind that the flashlight would illuminate objects by shining light through air, and never considered water. A good definition of normal performance cannot allow this. And if your alarm clock wakes you up again by falling on your head (why does it keep doing that? have you considered any alterations to your sleeping arrangement?) a would-be cheater might try to claim that it wakes you up normally by describing the designer's intentions too

broadly, observing that she intended that the clock would wake its user up at the appointed time by causing a sudden and intense sensation and that this is just what it did. Again, a good definition of normal performance cannot allow this.

I don't pursue what would probably seem like the most obvious strategy for dealing with the problem of the ambiguity of "the way": offering a scheme for specifying some privileged description, out of the range of possibilities, of *the* way in which a functionary is meant to perform its function. I think this approach is unlikely to work. As discussed above, a notion of success non-accidentally related to final-causal organization should allow *some* deviation from the paradigm. But once we don't specify every last detail of the paradigmatic manner of succeeding, and describe the way of succeeding in looser terms, it comes to seem a losing battle to try to find terms that exclude every thinkable weird or unrelated way in which success might occur by accident.

My alternative approach does not privilege any one description of the way in which a functionary is meant to function. I advance a notion of normal performance that "cares about" the whole range of possible descriptions. It requires first, that there be at *a* way in which the functionary is designed or adapted to perform which it did perform; and second, that to the extent that the functionary's performance *deviates* from the paradigm – doesn't fall under some descriptions of the way in which it was designed or adapted to perform – it deviates in an 'allowable' or 'non-abnormalizing' way.

To make this point more precise, we can introduce, as Millikan does, a notion of normal *conditions* for the performance of a teleo-function. My notion of normal conditions is drawn directly from the definition articulating my theory of functions. Something *o* is meant to do π for *X*'s purposes when *X* is *F* rather than *F'* because *X*'s being *F* is more conducive to *o*'s doing π than its

being F' , in some conditions C . Any conditions C that make such a function attribution true for some X, o, π are normal conditions for o 's performance of π , relative to X 's purposes.

On my approach here, there is not one distinction between normal and abnormal conditions, but rather a whole range of overlapping and nested regions of logical space each of which describes one respect in which conditions may be normal. Birds have hollow bones rather than solid ones because they're more conducive to flying under gravitational and atmospheric conditions within such-and-such a range. They also have hollow bones rather than solid ones because they're more conducive to flight on Earth. If a bird found herself transported to another planet with similar atmosphere and gravity, she would be in normal conditions for flying in one respect and in abnormal conditions in another respect. Likewise, if the strength of Earth's gravity suddenly increases tenfold, and makes flight impossible hollow bones or no hollow bones, then again, she is in normal conditions in one respect and abnormal conditions in another.

The spirit of what constitutes allowability here is that the performance can deviate from the paradigm as long as it only does so due to factors which do not introduce a disproportionate amount of luck in the explanation of success. More specifically, to the extent that the functionary performed its function in a way that departs from the paradigm, it must be that appealing to respects in which *conditions* depart from normal conditions satisfactorily answers the following question: *why did o perform π in a manner that departed from the paradigm in such-and-such ways, rather than either doing π in the paradigmatic way or not doing π at all?*

For example: place Wyoming in a bubble, fill it with Helium, and let its wolves communicate with one another with Donald Duck howls. In this situation, the wolves are in normal howling conditions under some broader descriptions of these conditions and outside of normal howling conditions under some narrower descriptions thereof (in particular those that mention *air*). They

also use howls to communicate in the normal way under broad descriptions, but the way they do so doesn't fit some narrower descriptions (particularly those that mention the howls propagating through *air*). To the extent that a performance is unusual in these more specific ways, this is precisely because the conditions of performance are abnormal in corresponding ways. But because the conditions were more broadly normal in the way that matter for success, the wolves still managed to use their howls to alert one another of their presence. In sum, the manner of communication deviates from the paradigm, but not in a way that introduces a disproportionate role for mere good luck.

Conversely, in the case of the butterfly-effect owl, characterizing the ways in which the owl's conditions did and didn't deviate from the paradigm won't explain why the feathers thermoregulated in a way that mismatched the paradigm rather than either better conforming to the paradigm or not thermoregulating at all. Sure, ways in which the conditions deviate from the paradigm might explain why the feathers didn't thermoregulate normally: it was too hot, so the normal warming effect of the feathers would be no help. But once we point this out, we can't see why the feathers did any thermos-regulating at all by merely pointing to the respects in which conditions were still normal (terrestrial atmospheric conditions, etc.). Rather, we need to cite some additional intervening factor – in this case some very complicated, butterfly-effect-y chain of atmospheric events – to see why the feathers performed their function in a way that departed from the paradigm rather than not performing their function at all.

This contrast suggests the following method for defining normal performance: something performs its function normally when *to the extent to which it is in normal conditions*, it performs in the way of performing to which its MPOs conduce in virtue of which it has those MPOs.

More precisely, we can define:

Definition (Performing normally): Let $OP^X\langle\pi(o)\rangle$ (i.e. let O be meant to do π for X 's purposes) and let it be the case that $\pi(o)$ (i.e. that o does π). o performs π normally relative to X 's purposes iff

- a) [*Normalcy in Some Respect*] There is at least one way of doing π , \bar{w} , which is a normal way for o to do π and in which o does do π ,⁴ and
- b) [*Proportionate, Explicable Departure*] For each way of doing π , w that is a normal way for O to do π to which O 's doing of π does not conform,⁵ there are some more determinate normal conditions C and some less determinate normal conditions C' for O 's doing π , such that O does π but not in w rather than doing- π -in- w -or-not-at-all because O is in C' but not in C

§3 – Toward a Notion of (Conditional) Requirement

A hypothetical imperative “if you’re doing Π you ought to ϕ ” expresses some relationship between the goals of the activity of doing Π with the activity of doing ϕ , one that “points strictly” from the pursuit of the goals to the doing of ϕ . Call this target notion that of *calling for* as in “doing Π calls for doing ϕ ” or ϕ -ing is called for by Π -ing.” The notion of being called-for will take some work to spell out precisely. Begin by distinguishing this notion from a natural first idea about when the expression of a hypothetical imperative is in order. This is the idea that hypothetical imperatives apply in conditions where Π -ing *alethically necessitates* π -ing. According to this idea, the hypothetical

⁴ That is: There exist a way of doing π , \bar{w} , and a feature of X , F , such that

- i. X has F as a mark of purposive organization around the purpose of O 's doing π (X has F rather than some F' because being F better conduces to such things as o doing π) in turn because X 's being F conduces to O 's doing π in \bar{w} , and
- ii. O does π in \bar{w} .

⁵ That is, for each way w of doing π such that

- i. X has F as a mark of purposive organization around the purpose of O 's doing π because X 's being F conduces to O 's doing π in w , but
- ii. O does not do π in w ,

imperative *if you're Π -ing you ought to π* is in order just when it is impossible to Π without π -ing. This condition is too demanding: consider cases where it is possible that an objective come about without the employment of some specific means, but only in case of a happy accident. *If you're trying not to take on any debt in the coming year, you should make a budget that allots expenses in keeping with your regular income* seems like a pretty good hypothetical imperative. But it is not *impossible* to avoid debt while living beyond your means since such happy accidents as lottery jackpots may occur.

I will suggest that a means is implicated by an end in the way that renders the expression of a hypothetical imperative appropriate, the implication is not alethic but teleological: it involves logical relations that hold not necessarily in all cases, but “as long as things work the way they are supposed to.” Simple ways of cashing out this idea of “teleological implication,” however, do not capture the intended notion.

The most straightforward way in which some u 's doing ϕ may be a teleologically, rather than alethically implied by some o 's doing π – of being, in some sense, a consequence of o 's doing π as long as things go as they're meant to – if o 's doing π *normally* entails that u does ϕ . Call this relation of implication *concomitance*.

The fact that ϕ is a concomitant of π is, too weak a condition for doing ϕ to be hypothetically imperative. This is because hypothetical imperatives, at least those of the sort that I mean to articulate, communicate that something is necessary as a *means to an end*, and ϕ s may be concomitants of π s not as means to them but as mere after-effects or side-effects. The accumulation of a waste product may be a concomitant of a metabolic process, but we don't want to construe hypothetical imperatives as in the business of commanding the accumulation of waste.

A stricter notion of teleological implication, one which accommodates the above point, can be defined by appeal to the notion of a *sole normal means*, that is, it is a concomitant that is specified to be a means to some operative plan: some μ such that (in general, or in a specific situation) the normal accomplishment of some π can only be a matter of doing $\mu \gg \pi$, of doing π by doing μ .

At first blush, there is some intuitive appeal to the idea that being a sole normal means is the kind of teleological implication that constitutes some π calling for some μ , for the espousal of a hypothetical imperative being in order. I do not think this suggestion survives scrutiny, however. In particular, I don't think being a sole normal means can be a necessary condition for being called-for, being a fitting subject of a hypothetical imperative. This is because hypothetical imperatives can properly be used to prescribe the *best* available means to normal performance, even when it is not the only available means. Think about what a golf instructor does: she imparts on her students various principles of swing mechanics, not because it is *only* swings that conform to these principles that ever non-accidentally produce good shots, but because conformity to this principles is the *best* way of producing good shots she knows. Though the notion of a sole normal means does not give rise to an appropriate notion of hypothetical imperativity straight away, sole normal means are very closely related to what I will call *demands*, which will be a crucial building block in my effort to define called-for-ness.

Having seen the problems with the straightforward attempts to define called-for-ness as a matter of implication of normal performance, I will take us down the path of a slightly different approach. This approach is abstractionist. The idea is to start by thinking about the actual process of some teacher or commander or dictator, or self-legislator, giving some U (perhaps itself) instructions about how to do Π . A choice of instruction may make sense for a specific instruction-giver based on her own idiosyncrasies – for instance, her personal preferences about how U is to do Π , or the kinds

of instructions she knows how to give. But some facts about what makes sense as a hypothetical imperative depend only on the nature of the task and the situation of the recipient of the instructions. I will think about the notion of hypothetical imperativity by thinking about just these “instructor-independent” constraints on what would count as fitting instruction in the performance of a task.

I think a little reflection reveals that something like the following three conditions must be met in order for ϕ -ing to be called for by Π -ing. There must be some point in ϕ -ing relevant to Π -ing: doing ϕ must further some purpose involved in doing Π . Second, the functionary doing Π must actually be able to do ϕ . Finally, the recipient must be able to do ϕ without facing some *conflict* with respect to doing Π , i.e. must have a way of ϕ -ing that doesn't threaten to frustrate other purposes involved in doing Π apart from the ones it furthers.

The goal of the next two sections is to spell out the idea that being called for, being conditionally required, being hypothetically imperative, is a matter of meeting these three conditions in naturalistic, becausal terms. To set up this framework, then, we need to think about what features of a circumstance and activity that a giver of instructions in the performance of a task properly responds to in producing those instructions (apart from their personal preferences or motivations). I will suggest that what an activity calls for emerges from the interplay, in the circumstances, of the range of possible demands that the recipient of the instructions might face in performing the task and the range of capacities, skills or strategies at that recipient's disposal in meeting those demands. I will call this range of possible demands a *remit* and the range of capabilities with which something may meet these demands a *repertoire*. Together, these notions make up a kind of “coordinate system” within which to specify a functionary's teleological predicament. To specify what demands they

might face and what responses they might produce to meet them is to specify their teleological situation.

These notions, along with the notion of an *effort*, which will be defined in terms of remit and repertoire, will provide the terms in which I can make good on the tripartite definition of called-for-ness sketched just above. It is to the definition of these three notions that I turn in the following section.

§4 – Remits, Repertoires, and Efforts

The first step in defining a remit, a set of potential demands on some entity, is to define the notion of an actual demand. The relevant notion of being demanded is the notion of an exclusive normal means, loosened up in two minor ways. First, allow that demands can hold not just between some plan μ and some operative plan π , but also between some plan μ and some broader encompassing activity Π which constitutively grounds the operativity of one or more plans π .

For instance, limping constitutively involves having the operative aim of locomoting, though the nature of the activity isn't exhausted by specifying this aim or any other. But because limping entails some purposes (operative aims), the exclusive normal means to those aims will thereby count as demands of limping. Likewise, *working on your health* probably constitutively entails, at various moments and in various circumstances, the operativity of various aims, but it is not clear that we could exhaustively characterize that activity by articulating a specific operative plan. And again, to the extent that in a given situation working on your health entails having some operative aim, the exclusive normal means to that aim count as a demand not only of that aim, but of working on your health in general.

The second tweak involves allowing that π itself is a demand of an operative plan π , even though it is not (at least arguably not) an exclusive normal means to itself, and further that when μ is an exclusive normal means to π , then $\mu \gg \pi$, doing π by doing μ is a demand of doing π as well, even though $\mu \gg \pi$ is not (or at least arguably not) a means (hence not an exclusive normal means) to doing π .

Demands: Let $\Pi(V)$ be some activity of some item V . If there exist some π, u such that, $OP^X[\pi(u)]$, (u being meant to do π relative to X), is a constituent element of V doing Π , and such that $\mu(o)$ is the exclusive normal means for $\pi(u)$ relative to X , (i.e. it follows from u 's doing π normally that it is by means of o doing μ $\mu(o) \gg \pi(u)$), then:

- 1) $DEM^X\langle\mu(o)|\Pi(V)\rangle$ (o doing μ is incumbent pursuant to V 's doing Π relative to X).
- 2) $DEM^X\langle\pi(u)|\Pi(V)\rangle$ (u doing π is incumbent pursuant to V 's doing Π , relative to X).
- 3) $DEM^X\langle\mu(o) \gg \pi(u)|\Pi(V)\rangle$ (u doing π by o doing μ is incumbent pursuant to V 's doing Π , relative to X).

Note that it follows from this definition that, for any operative plan $\pi(u)$ beneficiary X and exclusive normal means $\mu(o)$, that $\mu(o)$ is incumbent pursuant to $\pi(u)$, (write “ $DEM^X\langle\mu(o)|\pi(u)\rangle$ ”), $\pi(u)$ is incumbent pursuant to itself (write “ $DEM^X\langle\pi(u)|\pi(u)\rangle$ ”) and that $\mu(o) \gg \pi(u)$ is incumbent pursuant to $\pi(u)$ (write “ $INC^X\langle\mu(o) \gg \pi(u)|\pi(u)\rangle$ ”).

For the most part, in what follows, I will omit reference to the beneficiary, for concision, because the beneficiary with which we are concerned will generally be the functionary itself or the organism of which the functionary is apart.

4.1 – Demands as Generalization of Operative Purposes

We can think of a plan's being demanded as a kind of weakened or generalized version of its being operative. By definition, when a plan is operative (as a proper function), it is also demanded. But plans of activity which are part of the normal functioning of a creature, but not quite operative

in the CCET sense may also be demanded. One rubric for such a situation is a case in which π is operative, and the exclusive normal means to doing π is μ , but there has been no selection for doing π by doing μ *as opposed to* doing π in some other way.

For example, suppose that giraffes' height is a product of selection for height, in turn as a way of selection for the ability to reach high-hanging leaves. Then a newborn giraffe, Greta will have a proper function of growing taller than 4m. Calling this function "*grow*," we can write $OP^{Greta}(grow(Greta))$. Additionally (and relatedly, of course) Greta will be meant to *reach leaves* hanging higher than 4m. Calling this function of reaching *reach*, we have $OP^{Greta}(reach(Greta))$. If Greta goes on to reach leaves hanging this high *normally*, it will be by growing higher than 4m: for Greta, relative to her own purposes, *grow* is the exclusive normal means for *reach*. But have giraffes been selected specifically for reaching leaves hanging higher than 4m *by* growing taller than 4m? It is far from clear that we should say that they have been. Mother nature need not have had any preference for giraffes that reached leaves higher than 4m specifically by growing to that height rather than, say, reaching these leaves by knocking the trees down that held them. We might imagine that for all we know, no giraffe capable of reaching leaves higher than 4m in that way ever materialized, and so mother nature never had a chance to select against giraffes who got their food by this alternative means, nor *would* it have selected against them, had it been given that chance. Given these considerations, it may well be argued that, though Greta has some genes rather than others because of their comparative conduciveness to accomplishing *grow* and she has some genes rather than others because of their comparative conduciveness to her accomplishing *reach*, she does not have any genes rather than others because of their comparative conduciveness to her accomplishing *grow* \gg *reach*. Nonetheless, because the one is the exclusive

normal means to the other, we can say that *grow* \gg *reach* is demanded of Greta, that $DEM\langle grow \gg reach(Greta)\rangle$.

There is one more dimension along which demands are a kind of generalization of operative plans. This dimension has to do with specific contingencies of a contingency plan. Suppose a creature is organized around some contingency plan and a specific contingency comes up, but the creature doesn't (or cannot) adapt to the contingency so as to perform the corresponding component plan. Then the component plan corresponding to that constituency will not be operative in the CCET sense, but it will be demanded, and in particular, it will be demanded by the contingency plan.

For example, let Brad be a scout bee and let him be governed by the operative plan that makes the production of a specific dance contingent on the direction in which nectar lies relative to the hive. We can represent this as $OP^{Brad}\langle [P_\theta \mapsto \delta_\theta](Brad)\rangle$. Suppose that Brad encounters nectar at a heading of 104, but that for whatever reason, he doesn't adapt to the relevant "setting," even partially. In such a case, because Brad's physiology hasn't adapted to the circumstance, he is not organized around doing δ_{104} – it is not the case that $OP^{Brad}[\delta_{104}(Brad)]$. *But* there is nonetheless a sense in which we want to say that Brad is here meant to do δ_{104} . And here the notion of a demand comes in. Because, though its settings haven't been tuned, Brad still possesses the relevant dancemaking equipment, and therefore, the *contingency* plan $[P_\theta \mapsto \delta_\theta(Brad)]$ is still operative for him. And *in this particular circumstance*, in which he encountered the nectar at the particular heading he did, there is just one way he could execute this operative plan: this would involve Brad doing $[P_\theta \mapsto \delta_\theta]$ by doing δ_{104} . Because this is the only means by which Brad can execute an operative plan, it is, *a fortiori* the only normal means by which he can do so. It is therefore demanded of Brad: $DEM^{Brad}\langle \delta_{104}(Brad) | [P_\theta \mapsto \delta_\theta](Brad)\rangle$.

4.2 – Discharging Demands Normally

I have just said above that the notion of demand is a kind of generalization of the notion of an operative plan: plans connecting operative plans and their necessary means need to be operative but they are demands; and the actualized branches of an operative contingency plan need not be operative, but they are demanded. The notion of normal performance has been defined to apply to operative plans; but we may readily generalize it so that we may speak of the normal performance in keeping with some plan whether that plan is operative or merely a demand of some operative plan.

We can derive this generalization directly from the ways in which the notion of a demand generalizes the notion of an operative plan. When we have a demanded plan, by moving from actualized branch to overall contingency plan, or by moving from means (or means-end conjunction) to end, or by stringing together some sequence of these moves, we get eventually to an operative plan pursuant to which it is demanded. So to get a notion of normal performance of some demanded plan, we just need a way of defining normal performance of a contingency in terms of normal performance of an operative plan, normal performance of a an end-by-a-means in terms of normal performance of the end, and of normal performance of means to some end in terms of normal performance of the end.

The first two cases are simple because in each case, performing the demanded plan amounts, necessarily, to performing the relevant operative plan. For instance, doing π_3 in the circumstance P_3 constitutes doing $\kappa = [P_1 \mapsto \pi_1, P_2 \mapsto \pi_2, P_3 \mapsto \pi_3]$; indeed it is the only way of doing κ . And doing $\mu \gg \pi$ constitutes doing π . So we can define normalcy in each of these cases simply as a matter of the constituted performance being normal. A functionary performs some demanded branch π_3 normally just in case, in so doing, it performs κ , the contingency plan normally. Likewise, a functionary performs $\mu \gg \pi$ normally just in case in so doing, it performs π normally.

In the case where a means μ is demanded by some operative end π but the performance of the means is not such that it will necessarily amount to performance of the end, then we can take the basic idea of the definition of normalcy and measure the way in which μ was done against the information we get about how μ is to be done from the ways in which the relevant beneficiary, X is adapted for doing π . So o performs μ is normally just in case

- 1) There is some way \bar{w} of performing π that the beneficiary, X is adapted to conduce to, where doing π in \bar{w} is compatible with doing it by means of o doing μ in exactly the way it did.
- 2) For each way w in which X is adapted to doing π such that doing π in w involves doing μ in some way that *differs* from how o in fact did μ , there are some more determinate normal circumstances for doing π , C and some less determinate ones C', such that it is true that the o did μ happened but didn't do it in a manner consistent with thereby performing π in w , rather than performing μ in such a manner or not at all, because it was in C' but not in C.

4.3 – Remits

The notion of a demand is directly related to that of what I call a *remit*: the range of demands that may *possibly* arise for a functionary. Continuing the previous example: the idea is that prior to any specific circumstances in which a need arises, Brad may be governed by the operative contingency plan [$P_\theta \mapsto \delta_\theta$], but there won't be demanded of him the completion of any specific plan dictating the production of a specific dance, like δ_{71} . But insofar as it is possible for this demand to arise, it is a potential demand in his *remit*. More generally, specific arbitrary determinations of some operative or demanded plan are in a remit, even if these determinations are new. Suppose a frog has adaptations for catching bugs and there is some species of bug that has not yet, but could, encroach on its environment (and is, let's say, roughly as nutritious and as catchable as the bugs the frog is "used to"). Catching these bugs is in the frog's remit because it is a potential demand.

Potential demands are important as a contribution to called-for-ness: a specific question one that arises in instructing the functionary in the performance of a task, one that arises from the nature and circumstances of the task itself and not from any peculiarities of the instructor, is the question of what it is possible that accomplishing the task will require. A way in which some activity ϕ -ing might fail to be called-for by an activity A is to be *pointless*: it will take a little finessing to spell out what it means for there to be no point in doing ϕ , as far as satisfying the aims involved in A is concerned. But a first approximation is that there is nothing that the activity A could turn out to require – could be a potential demand thereof, which the doing of ϕ would further.

4.4 – Repertoires

The remit of a functionary constitutes the range of challenges that an entity might face under the purposes it is organized around. A sort of counterpart to this notion is that of a functionary's *repertoire*, which is the range of activities at its disposal by which it might act in keeping with those demands, if and when they arise.

To define the repertoire of a functionary, we need to establish what it means for there to be prospects for the performance of a function.

Prospects Say that o has *prospects* of performing some π in its remit iff it is possible for o to perform π normally.⁶ More generally, for some π in o 's remit say that there are prospects for

⁶ Strictly speaking, the notion of prospects is relativized to a locus of purposive organization. Consider again the archer and the nanobots. Relative to the purposive organization of the archer herself – of her training, of the psychological results of her training, and of her movements in firing the arrow – the prospects are *nil* for hitting the target normally. But in it can hit the target normally with respect to the purposive organization of the nanobot system as operated by the archery instructor, and so has prospects relative to that locus. However, this detail will not be centrally important in the ensuing discussion, because the target claim will always involve just one sort of locus of purposive organization; it will deal with the functions of parts aspects or behaviors of people as organic functions, in relation to that person themselves as locus.

Moreover, because the performance of some demanded plan μ may in principle be normal pursuant to some operative plan by which μ is demanded and not normal with respect to some other plan by which it is demanded, it is possible for some functionary o to have prospects for doing μ pursuant to one operative plan and to lack prospects pursuant to another. So, for instance, suppose that someone who doesn't know about the nanobots has the aim of using a series of

doing $\rho \gg \pi$ (whether or not ρ itself is in o 's remit) just in case it is possible for o 's doing $\rho \gg \pi$ to constitute o doing π normally.

A repertoire for performing some demanded plan may be defined as follows:

Repertoire: Let π be some plan in o 's remit pursuant to ϵ relative to : $REMIT\langle\pi(o)|\epsilon(o)\rangle$. Then say that ρ is in o 's repertoire for doing π ⁷ iff there exists some π' such that

- 1) $\rho \gg \pi$ is a potential demand in o 's remit and
- 2) o has prospects of doing ρ .⁸

To condense this notationally, we can write $REP\langle\rho(o)|\pi(o)\rangle$.⁹

The notion of repertoire here goes beyond what is perhaps the most natural interpretation of the natural-language term. Things that you cannot now do but could learn to do in the course of normal functioning count as in your repertoire, even though we might not be inclined to use the word “repertoire” this way ordinarily. Say that some activity ρ is in your *immediate repertoire* for doing π if it is in your repertoire for doing π and you have prospects of doing it right away. An immediate repertoire, as I have just defined it, corresponds more closely to the ordinary notion of repertoire.

It is noteworthy, and important, that some ρ may be in some functionary's repertoire for meeting a demand π which is impossible. One important point of saying “ o 's doing this is in its repertoire for doing that” is that it informs us that “doing that” is a candidate for the point or aim of

ten shots to confirm to himself that the archer is a capable archer (he can't observe the archer firing – he can only see where her shots land). And suppose at the same time that the archery instructor is using the same series of shots to confirm that she is capable of using the nanobot network to direct shots to the bullseyes. In such a case, the shots have prospects of hitting the bullseye in a way that is normal for confirming to the instructor that he can operate the nanobots, but they do not have prospects of hitting the bullseye in a way that is normal for confirming the archer's capability to the naïve observer. This detail will be unimportant because the key clauses of the notion of called-for-ness will involve the existence of prospects for plans pursuant to some operative plan, without regard to whether there are prospects for this same plan relative every operative plan by which it is demanded.

⁷ Relative to X , pursuant to ϵ .

⁸ Relative to X , pursuant to ϵ .

⁹ Adding notational parameters sufficient to convey the most general case, we can write $REP^X\langle\rho(o)|\pi(u)|\epsilon(v)\rangle$ for ρ is in o 's repertoire for getting u to do π pursuant to v 's doing ϵ .

“doing this” on any particular occasion that θ does it. And it is possible to do something with the aim of doing something else that you cannot actually thereby do. If we don’t allow that an activity can be in a functionary’s repertoire for doing something that this repertoire is inadequate to allow the functionary to do, we will be illicitly restricting the range of possible candidates for what some functionary might be classified as “up to,” assuming, effectively, that it must be “up to” something that it is actually capable of.

For example: suppose that a calculator is programmed to compute square-roots. Suppose that it is programmed to do this by taking a decimal digit input and performing successive approximations – out to one decimal digit, then to two digits, and so on – until the approximation becomes perfect – if it ever does. Such a calculator cannot compute the exact square-root of two, since it is an irrational number whose decimal digit representation goes on forever. But computing the exact square root of 2 might be demanded of it – if we punch in ‘2’ and press the square root button. In such a case, the machine won’t be able to meet this impossible demand, but we will be able to say, in an informative way, what means it would have to take in order to compute the exact square root of 2 normally, in keeping with its organization around square-root-computing. Specifically, we can say that it would have to compute the square root by computing a finite sequence of approximations – to one digit, to two digits, to three digits – until it reached the exact value. And though it is impossible for the machine to get the exact value by this method, it is possible for the machine to take the steps involved in generating this sequence. Each of these steps is suited to play the role of an interposed π' - a step towards meeting the demand, which it can take, but which cannot culminate in final success. The means by which the calculator achieves these estimates are therefore part of its repertoire for computing the square-root of 2, a repertoire inadequate to the demand.

A functionary's remit, as well as its repertoire, is relative to some beneficiary, some locus of purposive organization, because the question what counts as a normal performance – and hence what means a normal performance might require – depends upon specifying a beneficiary whose marks of purposive organization would have to satisfy the relevant explanatory conditions constitutive of a normal performance. So for instance, imagine a bacterium that cannot metabolize fructose in a laboratory set up to genetically engineer that bacterium to metabolize fructose. It is not in what we might call the *proper* remit of the bacterium – the set of possible demands implicit in the purposive organization of the bacterium itself – to metabolize fructose. But it is in the bacterium's remit, relative to the laboratory, to metabolize fructose: if it is successfully engineered and then placed in a fructose-rich environment, then it may turn out that the only means by which it can perform various metabolic functions is to metabolize fructose. And the possible future in which the bacterium acquires this ability and the need to exercise it arises, and it discharges this demand, would be *non-accidental* or normal relative to the purposive organization already in the laboratory. As usual, unless otherwise noted, talk of an entity's remit will be talk of its proper remit, what might be demanded by the normal performance of the ends around which it itself is organized.

4.4 – Efforts

In one sense, it is very straightforward what it is the function of producing a hypothetical imperative to bring about. “If you're Π -ing, you ought to π ” prescribes π -ing-if- Π -ing. But there are cases in which the production and consumption of an ought-claim goes off without a hitch, and in initial steps, everything goes according to plan, but in which these steps ultimately come to nothing. If you're trying to survive alone in the wild, you should catch small animals to eat” is perhaps a reasonable hypothetical imperative, but someone's reminding themselves of it and acting competently in light of so doing need not end up catching any animals. In such a case, the use of the

imperative doesn't do what it is more distally meant to do, but it does its more proximate job: its recipient, we'd naturally say, makes an *effort* to do what the imperative prescribes. If we are understanding called-for-ness as a matter of there being something like a "job opportunity" for something that behaves like an ought-symbol, the question of what it is for such an opportunity for apt functioning involves, should be sensitive to not just the possible and probable *distal* effects of the use of the ought-message, but also to its possible and probable *proximate* effects.

The question of whether the demands and ability of a functionary o make it so that there is a job for an ought-to- ϕ -message, sent to o , to do, is thus fundamentally a question of whether given the challenges and abilities that arise in o 's predicament, there is occasion for the use of an ought-to- ϕ message in making for a certain kind of effort of o to do ϕ . That a focus on the proximate job of an ought-message, its production of an effort to do what the ought-message prescribes, is necessary in thinking about what it is for ϕ -ing to be called-for, for there to be a job for an ought-to- ϕ message to do, is particularly pronounced if we turn our attention to ought-to-*be* messages. "You ought to feel sympathy for the bereaved" prescribes feeling sympathy for the bereaved; but there can be no mistaking that the proximate response of someone who accepts such a prescription when reminded of it may not result in feeling sympathy for the bereaved just like that – acting on it is in the first instance a matter of making an effort to feel sympathy for the bereaved. This is important because an imperative to believe the truth is also a kind of ought-to-be, so the fact that its proper functioning results in the first instance in an effort to believe the truth is similarly salient.

We thus need a notion of effort, of some activity's being specifically directed at some result, in order to define called-for-ness, the circumstance in which there is a job opportunity for an ought-message. Here is how I propose to do so. Start with the idea that when some α is in a repertoire for doing some π this articulates that π is a candidate for what o might be up to in doing α . But it need

not be the exclusive candidate – doing α may be at o 's disposal as a way of potentially doing various things. At the same time, requiring that doing α be only in o 's repertoire for doing π , and not for any other μ is too restrictive: if doing α is an effort by o specifically directed at doing π can also be a way for o to do μ – the point of doing α may be to do μ as a *means* to doing π or it may be to do π as a means to do μ , or the point may be to do π and μ at once. What we want to avoid is a case in which π is a possible point of doing α in a way that makes it optional, and where other tasks for which α is in o 's repertoire are *alternative* points of π . So for example, a heart's contractions constitute an effort to circulate blood because whatever it might be called upon to do by means of these contractions, it is something it may be called upon to do by circulating blood, as a means to circulating blood, or while circulating blood. There is nothing, in the course of a normal animal life, that the heart may be called on to do by means of its contractions *instead* of circulating blood.

Effort:

o 's doing α is an *effort* to do π just in case α is in o 's repertoire for doing π , and for any μ such that α is in o 's repertoire for doing μ , either

- 1) α is in o 's repertoire for doing $\mu \gg \pi$,
- 2) α is in o 's repertoire for doing $\pi \gg \mu$, or
- 3) α is in o 's repertoire for doing $\pi \& \mu$.

The idea of this definition is that when α -ing is an effort to π , there is some respect in which its doing α cannot be fully successful, or fully capitalized upon, unless it results in the doing of π . All roads of characteristic activity that extend from α lead through π : if one leads to any other μ , they lead to μ on the way to π (clause 1) or to π on the way to μ (clause 2), or it leads to μ and π simultaneously (clause 3). Activities from which multiple roads of characteristic activity diverge, leading alternatively to π or to μ are not specifically directed at doing π , and are not efforts.

§5 – A Recipe for Ought

In the previous section, I introduced notions that characterize the predicament of some thing with a purpose – its remit, the possible demands its ends may place on it, and its repertoire, the ways in which it is capable of meeting those demands – as well as the notion of an effort, of some activities requiring a certain result to be fully successful or capitalized upon. These notions allow us to precisify the three conditions I mentioned that make up called-for-ness: prospects, a point, and a lack of conflict.

First: there must be prospects for σ of doing ϕ , or else an effort to do ϕ would be simply hopeless. This is taken care of with the definition of prospects above. The second and third conditions – that there be some point to doing ϕ and that doing ϕ not face a conflict – will take some more effort to work out.

5.1 – Having a Point

Note that what matters, as far as ϕ -ing's having a point with respect to doing Π goes, is that doing ϕ furthers some *ultimate* as opposed to an *intermediate* aim of doing ϕ . Suppose your goal is to ring the bell and flexing your arm is a way to pull on the rope to which the bell is attached, but the rope is hanging by a single fiber and will break if you pull on that. Then is there a point to flexing your arm, as far as ringing the bell goes? Well, it can successfully achieve one aim: you do have prospects of pulling the rope by flexing your arm. But even though we can point to something that flexing your arm is good for *vis-à-vis* ringing the bell – pulling on the rope – flexing your arm is still ultimately pointless in your endeavors to ring the bell, which we see when we take a step back to look at a more distal aim. Articulating what is required for ϕ -ing not to be pointless is easy if we're

considering a hypothetical imperative to ϕ if engaged in a specific goal-oriented activity like ringing the bell: we just care about whether doing ϕ improves your prospects of ringing the bell. It is harder if the activity in question is something that involves aims, but for which there is no clear way to articulate some specific set of one or a few ultimate aims, like playing the bassoon, or working on your health, or living your life.

To handle the idea of doing ϕ 's having a point in a way that handles all cases, we need to pursue a negative route: when some act ϕ would be pointless with respect to the aims of some activity Π , we would have to find – eventually, when we zoom out at some level of distality of aims – that any possible purpose to which doing ϕ could putatively be put in the course of doing Π would be futile. So the idea will be that, no matter how you try to do this, you will not get a set of aims none of which have prospects.

To get this right, we will need the notion of a *subordinating set*.

Subordinating Set: Let ϕ be in the repertoire of o for doing Π , and let $S = \{\pi_1, \pi_2, \dots\}$ be a set of plans in o 's remit for Π . S is a subordinating set for $\phi(o)$ with respect to Π just in case for any π' in o 's remit such that ρ is in o 's repertoire for doing π' , there exists some π_i in S such that either

- 1) π' is identical to π_i
- 2) $REP(\phi(o)|[\pi' \gg \pi_i](o))$, ϕ is in o 's repertoire for doing $\pi' \gg \pi_i$, or
- 3) $REP(\phi(o)|[\pi_i \gg \pi'](o))$, ϕ is in o 's repertoire for doing $\pi_i \gg \pi'$.

The idea is that a subordinating set for ϕ is one complete set of possible points of doing ϕ . Think of possible uses to which doing ϕ might be put in doing Π can be organized into a branching path: doing ϕ may lead to the achievement of some range of more proximate ends, each of which may in turn lead to any of several more distal ends. A subordinating set is a set of nodes in this branching

path such that whatever path you take, you eventually get to one of the select nodes, though perhaps sooner on some paths than others.

Though a subordinating set for ϕ represents a complete set of possible points of doing ϕ , there is plenty of room for choice in its composition: choices of whether to include more proximal or more distal possible points, of whether to cut off the branching paths that lead from ϕ either early on, or further down the road. The condition of pointfulness, is that no matter what subordinating set you pick – no matter how far out on the chain of effects its elements are – you will always have at least one element of the subordinating set for which doing ϕ improves the prospects.

Pointfulness: Say that there is a point of doing ϕ for o in doing Π ($POINT\langle\phi(o)|\Pi\rangle$) just in case for any subordinating set S of $\phi(o)$ with respect to Π , there exists some π_i in S such that $\phi(o)$ would improve o 's prospects for doing π_i .

5.2 – No Conflict

The idea that doing ϕ doesn't conflict with other aims of doing Π can be cashed in terms of making prospects no worse for the various demands in the remit for doing Π . Though the basic idea is simple, we need to be careful on three counts. First, is a minor detail: sometimes when doing ϕ worsens prospects for some π , this doesn't create a conflict because doing so wouldn't worsen any prospects downstream of π . Doing ϕ might close off π as an option for doing some π' because it pursues some other, better option. To address this condition, we can relax the requirement that doing ϕ not worsen the prospects of any π in the remit, to the condition that if doing ϕ worsens the prospects of doing any π , there is some subordinating set S for π such that doing ϕ does not worsen the prospects of doing any π_i in S .

The second wrinkle is that it is not important that it be the case that *any way whatsoever* for o to do ϕ wouldn't worsen prospects. What matters is whether there is *a* way for o to do ϕ that does not relevantly lower prospects. To accommodate this point, we can define a conflict-sensitive determinant.

Conflict-Sensitive Determinant: Let ϕ be in o 's repertoire for $\Pi(U)$. Then say that $\hat{\phi}$ is a *conflict-sensitive determinant* of ϕ for o pursuant to Π just in case

- 1) Doing $\hat{\phi}$ is a way of doing ϕ , and
- 2) For any π in the remit of o (or any other part of U) for Π such that o 's doing ϕ normally in doing $\hat{\phi}$ would worsen the prospects for π as compared to not doing ϕ there exists some subordinating set S of π with respect to Π such that o 's doing $\hat{\phi}$ normally in doing ϕ would not worsen the prospects for any π_i in S .

The third wrinkle has to do with the fact that the immediate effect of a ϕ -instruction is an effort to ϕ , which might or might not succeed. If we care about conflicts, we care about not just the conflicts that might arise from successfully performing ϕ , but also those that might arise from the effort itself. And here, we care about the conflict-sensitivity of the effort involved, not just the conflict-sensitivity of actually ϕ -ing; it's not good enough that doing ϕ in a way that is sensitive to conflicts is possible, if the efforts that could lead to so doing have a chance of failing and thereby conflicting with other ends of the governing activity. So we should want the effort itself to be performable in a conflict sensitive way. But this might appear just to push the issue back, starting us on a regress. Suppose we can make some effort which is itself conflict-sensitive, to do ϕ in a conflict-sensitive way: what about the possibility of trying and failing to do whatever amounts to this conflict sensitive effort, and thereby inviting a conflict?

Fortunately, I think the intuitions that drive the whole functional-representationalist idea about the notion of requirement provide a natural candidate for a "basic effort" to ϕ , where conflicts caused by making this effort do make a difference to the called-for-ness of ϕ -ing but conflicts which

would be caused by a failed effort to make this effort do not matter. This is the effort to get o to ϕ which the sending of the ϕ -instruction would itself constitute. If there is no way to get the message across to o to do ϕ that would itself avoid incurring any conflict, that is a problem for called-for-ness. But if the problem is merely that it is possible to try and fail to get the message across to o , in such a conflict-free way, and that this failure would itself invite conflict, that in itself creates no ambiguity about whether doing ϕ is called-for: it creates a problem at most, for attempting to instruct o to do ϕ .

5.3 – Being Called-For

We can put the strategies we've formulated to articulate the three conditions on called-for-ness – that there be prospects, that there be a point, and that there be a way of avoiding conflict – as follows. Note that the notion of an indicative intentional signal deployed in this definition will be defined in the following section.

Called-For-ness Say that o doing ϕ is called for by $\Pi(U)$ iff there is some signalling activity, call it σ_ϕ that serves as an indicative intentional icon instructing doing ϕ , and

(No Conflict) there are conflict-sensitive determinations of ϕ and σ_ϕ , $\hat{\phi}$ and $\hat{\sigma}_\phi$, such that

- 1) **(Prospects)** The plan, $\hat{\sigma}_\phi(_, o) \gg \hat{\phi}(o)$, that is, o doing $\hat{\phi}$ by means of a signal $\hat{\sigma}_\phi$ sent from some element of U to o , is in U 's repertoire for doing Π and
- 2) **(Point)** $POINT\langle U | \hat{\sigma}_\phi(_, o) \gg \hat{\phi}(o) | \Pi \rangle$, there is a point in o 's doing $\hat{\phi}$ by way of being instructed by the signal $\hat{\sigma}_\phi$, pursuant to doing Π .

§6 – Belief as Preparedness to Signal

My account of belief proceeds in two steps. First, I formulate a notion of an intentional signal: a tract of activity by a functionary that has the job of indicating that things are a certain way

or instructing an effort to bring about a certain circumstance. Then I introduce a notion of preparedness, that “teleologically enriches” the notion of a disposition and present an account of beliefs as conditions of preparedness to give an indicative signal in relevantly pertinent circumstances.

6.1 – Against Competition between Accounts of Intentional Phenomena

There are many teleo-semantic accounts of intentional states and episodes, so it is worth making some remarks on what I aim to accomplish in adding another.

The first is that, when thinking about the relationship between the various proposals here, I don’t think that competition is the right model. I come to reject this model by starting with an ontological point that I mentioned briefly in the preceding chapter. There, to try to push a separate point, I pointed out that all of terrestrial life is – because it is causally and spatiotemporally connected in a tree of common descent,¹⁰ a *single concrete particular* in an important and natural sense. What does this have to do with whether teleosemantic accounts of intentional states and episodes are to be pitted against one another in competition? Well, consider what this point about terrestrial life suggests about the question about how to define life.

A definition would – as definitions generally do for their *definienda* – tell us which properties of the life we know about, the life on Earth, are constitutive of its being life, as such (or follow necessarily from those constitutive properties), and which properties are idiosyncrasies that life elsewhere may not share. Likewise, such a definition would draw a distinction between different sorts of (actual or possible) phenomena elsewhere in the universe that are life, though they differ

¹⁰ More precisely, a bush of common descent and interorganismal exchange of genetic information (mostly, in modern times, between bacteria) (Haig 2020).

from terrestrial life in this or that way, and those phenomena that are not life, though they may resemble terrestrial life in some way.

What properties are we to abstract from the phenomenon of terrestrial life in order to specify this property? In an essay arguing against the idea that “life” should have a definition, Edouard Machery (2012) draws our attention to the point that different areas of inquiry would seem to be interested in different notions. For instance, virtual life researchers, who are concerned with the possible computer activity that is importantly like life, care about some ways in which known actual life might be like something a computer might do and uninterested in certain ways in which the two will not be alike. But a *synthetic* life researcher, who is interested in the possibility of synthesizing living things from non-living parts in a lab, would have their own kinds of similarities between known life and possible laboratory projects that are of interest to them, and possible differences which are not their concern. And the way these two projects sort between properties of known life they care about capturing and properties of known life that they are to deviate from is likely to differ. For instance, an artificial life researcher will likely be interested in virtual entities, constituted by computer activity, that evolve – and certainly with those that reproduce - whereas a synthetic life researcher might count it as a resounding success if they make something that replicates most of the biochemistry a modern bacterium, even if it dies before reproducing at all.

Pursuing a definition of life, Machery concludes, misunderstands the role that thinking about life plays – or better the plurality of roles such thinking plays – in scientific inquiry. I think that Machery’s observation about the ill-fit between the role of thinking about life and the idea of a single, once-and-for-all definition thereof can be explained, and his case strengthened, by noting, as I did above, that known life is in an important sense a single concrete particular. Known life is distinctive – different from the other phenomena we’ve encountered – in various interesting ways.

But why think that the best way to think about this distinctiveness is to try to formulate a definition that captures all and only the essential or fundamental elements of this distinctiveness?

If you think of a person you know well, you can probably enumerate a number of interesting and distinctive features of theirs. Do we think that there is a specific privileged subset of these that form the single, correct definition of the *kind of person* they are? Of course not. You might meet someone who has some of the distinctive features of your friend but lacks others. Are they the same kind of person? I think it's pretty clear here that all there is to say is that in some ways, they are the same kind of person, and in other ways, they are not. Likewise the life that we know of is a single occurrence, with various interesting distinctive properties. Some things that we can hypothesize – some of which we might eventually encounter – will be like it in some distinctive ways and unlike it in others. Others will be like and unlike it in different ways. To ask what boundaries define *the kind* of this single occurrence really is misunderstands the state of play – it is many kinds of things, just like your friend is many kinds of person.

Much the same thing can be said about the intentional phenomena of which we know. It can hardly be doubted that prior to any specialized knowledge in psychology or cognitive science or linguistics or philosophy of language or mind, the mental model by which we understand intentional phenomena are the states and doings of people: what people say, what they intentionally do, what they believe, what they intend, what they see and hear, and so on. We notice – or hypothesize – that various other things – non-human animals, cognitive processes, computers - exhibit states and activities that are like these in interesting and distinctive ways.¹¹

¹¹ What's more, we notice that some of the base or paradigm cases are like each other in distinctive ways; there seems to be, for instance some significant analogy between seeing-that and saying-that, as the effortlessness with which we comprehend metaphorical talk about "the testimony of the senses" attests.

But again, should we think that the way to understand phenomena of human thought – a particular occurrence (scattered across human history) with various distinctive properties – is by selecting some privileged set of distinctive properties of these phenomena to amount to count as their definitions? And consequently, should we think that these definitions yield singular unequivocal answers, when faced with activities or conditions – of an animal, of a brain or its parts, or of a cell a cell – that are like the phenomena of human thought in some ways and unlike them in others, that there should be a single unequivocal answer as to whether they are the same kind of phenomena. The thing to say is that the phenomena are like human thought in some ways and unlike them in others. Different definitions will capture different affinities.

I am by no means interested in articulating conditions necessary and sufficient for minimal intentionality. Nor do I think that the notion of intentional signal I will advance captures every distinctive feature of conscious human imperative or indicative thought. I simply want to present a notion that captures enough of the teleo-functional profile of the phenomenon to support an argument for Truth-Normativity.

6.2 – Contingency Sets and Acting-Like

The basic idea behind the notion of intentional signal I will introduce is that they have the job of organizing or sorting behavior of their recipients. Indicative intentional signals are efforts to organize the behavior of their recipients such that, whenever there are demanded of the recipient contingency plans in which P is a contingency, the recipient pursues the branch of the contingency plan corresponding to P. In this sense they are efforts directed at their recipients acting like P. Imperative intentional signals are efforts to get their recipients to do some ϕ . In this sense, they are efforts directed at their recipients acting like they're doing ϕ , in the most straightforward sense, *viz* doing ϕ .

Contingency Set

For any functionary o , any locus X , $CONT^X(o, P)$, the *contingency set* for o, P , is the set of all plans κ in the remit of o where κ is of the form $\kappa = [P \mapsto \alpha_P, \dots]$, i.e. the set of all contingency plans in o 's remit in which P is a contingency.

Next, define acting like P as doing α_P , the strand of the contingency plan corresponding to P , in any contingency plan κ in the contingency set for o, P , where κ is demanded, and where o actually has prospects of doing α_P in putative fulfillment of $\kappa(o)$.

Acting-Like

Say that o *acts like* P , (write $A_P(o)$) just in case for every κ where

- 1) κ is in $CONT(o, P)$, the contingency set for o, P
- 2) such that α_P , the branch of κ corresponding to the contingency P is in o 's repertoire for doing κ , and
- 3) κ is demanded of o ,

o does α_P .

Note that it is possible for a contingency plan $\kappa(o)$ to demand α_P , some branch corresponding to some contingency P , and even for o to have prospects for doing α_P , even if P is false or impossible. It is therefore possible for o to act like P is the case, by performing the relevant branches α_P normally, even when P is false, or even impossible. An example of this sort of dynamic occurs when Brad the scout bee finds nectar in Buzzton, but because of a neural defect or freak accident, his brain gets “set” to perform the “nectar-in-Beeattle” dance instead. The facts that he encountered nectar in Buzzton and the setting of Brad’s brain – corresponding to the production of the “nectar-in-Beeattle” dance – make conflicting demands of Brad. But because it is possible for Brad to perform the “nectar-in-Beeattle” dance normally, thereby discharging one of these conflicting demands, it is possible for him to act like there is nectar in Beeattle.

6.3 – Interpretive Schemes

My account of intentional signals will understand them as efforts by a sender to get a receiver to employ what I call an *interpretive scheme*, some pattern in the repertoire of the receiver for correlating messages from the sender with responses. An *indicative interpretive scheme* looks like this:

$$\begin{aligned} \Sigma_{\text{ind}}(U) = & \sigma_P(o_1) \mapsto A_P(o_2) \\ & \sigma_Q(o_1) \mapsto A_Q(o_2) \\ & \vdots \end{aligned}$$

This is a plan that dictates that o_2 respond in specific ways to different activities produced by o_1 – by acting like P is the case in response to σ_P , acting like Q in response to σ_Q , and so on. Call this sort of plan an *indicative interpretive scheme*. The idea is that, under Σ_{ind} , σ_P is to be interpreted by o_2 to mean P , and so on. Analogously, an *imperative interpretive scheme* Σ_{imp} is a plan of the form

$$\begin{aligned} \Sigma_{\text{imp}}(U) = & \sigma_\phi(o_1) \mapsto \phi(o_2) \\ & \sigma_\psi(o_1) \mapsto \psi(o_2) \\ & \vdots \end{aligned}$$

This is a plan that dictates that o_2 respond to certain activities σ_ϕ, σ_ψ , by doing ϕ, ψ , and so on respectively.

The basic idea behind my notion of an intentional signal is that the production of an intentional signal is an effort on the part of the sender to avail itself of an interpretive scheme that is in the receiver’s repertoire in a particular way. An interpretive scheme may be a more immediate part of a potential recipient’s repertoire, and occasioned in a wide, general range of circumstances.

Knowledge of a natural language involves possessing interpretive scheme in this sort of way. If you know German, you can respond to utterances according to their meaning in German right away and in all sorts of contexts, as the conventional meanings of German sentences dictate. An interpretive scheme might be immediately available to a potential recipient, but only occasioned in very specific circumstances, as when a sender improvises a signal the intended meaning of which relies on specific

details of the context, or when the sender and receiver make an agreement that institutes meanings for specific signals. An interpretive scheme may be in a recipient's repertoire but not its immediate repertoire, so that it would only be occasioned by some potentially lengthy enabling process. Natural languages a person doesn't know but could learn fall into this category.

6.4 – Intentional Signals Defined

An intentional signal is it is an attempt by a sender, to avail itself of some interpretive scheme in the receiver's repertoire, and in particular to get the receiver to do $A_P(\phi)$ by performing the corresponding σ_P (or σ_ϕ) in that interpretive scheme.

Intentional Signals

Consider some o_1, o_2 and some Σ_{ind} in the repertoire of o_2 , and suppose that in Σ_{ind} , $\sigma(o_1)$ maps to σ , suppose that o_1 does σ_P in an effort to get o_2 to do A_P by following $\Sigma_{ind}: EFF\langle\sigma(o_1)|[\Sigma_{ind} \gg A_P](o_2)\rangle$. Then o_1 's doing σ is an indicative intentional signal that P (Write $SIG\langle\sigma(o_1)|P, o_2 \rangle$).

For some o_1, o_2 and some Σ_{imp} in the repertoire of o_2 in which σ maps to ϕ , suppose that o_1 does σ in an effort to get o_2 to do A_P by following Σ_{imp} . Then σ is an imperative intentional signal meaning $\phi: o_1$, in doing σ , *instructs* o_2 to ϕ . (Write $INST\langle\sigma(o)|\phi(o_2)\rangle$).

This notion of an intentional signal is undemanding in the sense that it allows the sender of a signal not to be able to interpret it. The intentional content of episodes of conscious human thought and speech are importantly richer than this bare notion of intentional signal delivers on its own, because the producer of the thought or utterance will have the relevant interpretive scheme that they aim to elicit the use of by their audience in their *own* repertoire. This fact characterizes a dimension of awareness of what we mean by what we say and think that is not necessarily present in more rudimentary signalling activities that fit this definition. Since we are also potential consumers of the messages we produce, we possess a kind of "maker's knowledge" of them that distinguishes us as

sending messages rather than merely *emitting* them and does so without the overintellectualization that would be involved in supposing that each of our utterances is preceded by a thought of the form “I will say such-and-such words to communicate so-and-so.”

6.5 – Determinacy of Content: Preliminary Points

This notion of indicative and imperative signal allows a great deal of routine ambiguity in content. A scout bees dance is a notorious example of what Millikan calls a “pushmi-pullyu representation,” (2005) one that functions to coordinate a specific circumstance with a specific response. On the view described above, such representations count both as indicative signals of the kind of circumstance in which it is their producer’s function to produce them as well as imperative signals of the kind of response they are supposed to evoke in the recipient. So, for example, if Brad does dance δ_{108} , this counts as an indicative intentional signal indicating that nectar is present at a heading of 108. This is because such a dance is an effort to get the recipient, - the colony – to follow the interpretive scheme $[\delta_x(\text{Brad}) \mapsto A_{\text{nectar at heading } x}(\text{colony})]$ and to thereby act like there is nectar at heading 108. Any demands that Brad’s doing δ_{108} helps to fulfill in the normal course of activity will be some means to the colony acting like there is nectar at a heading of 108 by adhering to the above interpretive scheme; or it will be something the colony does *by* adhering to this form of interpretive activity; or it will be something it does or could do in addition to this interpretive activity. There is no strictly alternative – rather than instrumental, consequential, or complementary – way for use to be made of Brad’s performance in the course of normal functioning. It is an effort to get the colony to act as though there is nectar at heading 108 because it is directed at getting all of the contingency plans that are demanded of the colony and which involve the contingency of nectar at this heading to follow the branch corresponding to this contingency. These demanded contingency plans are – on the supposition that the function of δ_{108} is to produce a single uniform

type of response – just (perhaps multiple different, normally equivalent, ways of describing) the plan that correlates the direction of nectar with directions of flight of forager bees. It is an effort to send the colony down the 108 branch of this contingency, as any δ_x would be to send it down the branch corresponding to x . But this is just to say that the normal working of δ_{108} is to draw on the availability of the *imperative* interpretive scheme associating dances with specific flight responses, hence that it is an imperative signal instructing such a response.

But as I alluded to in passing above, the indeterminacy is not just a matter of potential indeterminacy as between indicative and imperative symbols. Notice that, because nectar is the bees' source of sugar, just as the bees are adapted to fly where the nectar is, they are adapted to fly where the usable sugar is. The liberality of the CCET notion of function is such that there is no requirement to choose a real function. But by the same token, this means that whenever one of these contingency plans is occasioned, so is the other (except perhaps in some extremely aberrant circumstances) and Brad's performance of his dance is equally an effort to produce both a flight in heading 108 as a performance in keeping with the contingency plan saying to go where the sugar is as to produce that flight as a performance in keeping with the contingency plan saying to go where the nectar is. Thus, it is inarticulate with respect to the distinction between these two contents – Brad's dance is a representation indicating that nectar is in a certain place as well as one indicating that sugar is in a certain place, rather than specifically being one or the other.

This is not to say that the content of the dance is indeterminate along every that the teleosemantic accounts have been accused of being indeterminate. Suppose for example, that there is some dupe chemical that scout bees mistake for sugar in the way that we might mistake sucralose or aspartame for sugar. It does not come out, on the account above, that it is indeterminate whether δ_{108} means that there's sugar at location 108 or that there's sugar-or-dupe at location 108. In order

for that meaning to be attributable to Brad's dance, it would have to have constituted an effort to have an interpreter use an interpretive scheme Σ that mapped δ_{108} to $A_{sugar-or-dupe\ at\ 108}$, a plan that involves acting like there is sugar-or-dupe at 108 in response to δ_{108} among other contingencies. And it can only be such an effort by Brad it is in Brad's remit: if eliciting the use of such an interpretive scheme is something that might be demanded of Brad in service of his fulfilling an operative plan normally. The plausible candidate for a such a demand would be the demand that he produce δ_{108} as an effort to elicit in the colony the use of some interpretive scheme like $[\delta_x(Brad) \mapsto A_{sugar-or-dupe\ at\ x}](colony)$, on the part of the colony.

But the elicitation of such an interpretive response is not in Brad's remit. This fact derives from the fact that acting in this way is not part of the remit of the colony. For doing $A_{sugar-or-dupe\ at\ 108}$ to be in the colony's remit, it must be well-defined: there must be some well-defined way in which the colony can act like there is sugar-or-dupe at 108. But for $A_{sugar-or-dupe\ at\ 108}$ to be defined for the colony, there would have to be contingency plans in the colony's remit in which there being sugar-or-dupe at 108 is a contingency. And it is not plausible that the colony is adapted to a contingency plan involving this disjunctive sort of contingency.

To be sure, there are contingency plans operative for the colony that involve something normally coextensive to there being sugar-or-dupe at 108 as a contingency: for instance $\kappa_{sugar} = [sugar\ at\ x \mapsto scouts\ go\ to\ x]$ is operative for the colony, and when the sugar is at location 108, it demands the corresponding branch. The colony has certain features rather than others because those features better conduce to the colony's behaving in accordance with this contingency plan than the alternative. But the contingency plan $\kappa_{sugar-or-dupe} = [sugar - or - dupe\ at\ x \mapsto scouts\ go\ to\ x]$ is *not* operative for the colony. Greater comparative conduciveness specifically to this sort of pattern of behavior – the pattern sending scouts in a particular direction in

specifically in different disjunctively described circumstances – is not plausibly an explanation of why some colonies outcompeted others: such a putative explanation adds an extraneous disjunct. Nor does $\kappa_{sugar-or-dupe}$ arise as a *demand* on the colony as an exclusive normal means to some operative plan like κ_{sugar} or κ_{nectar} . True, when the colony conforms to κ_{sugar} normally, it also conforms to $\kappa_{sugar-or-dupe}$, i.e. the latter is a normal *concomitant* of the former. But in conforming to κ_{sugar} , the colony does not do so *by means of* conforming to $\kappa_{sugar-or-dupe}$, just as I don't slip on ice by slipping on ice-or-a-banana-peel. $\kappa_{sugar-or-dupe}$ is therefore not an exclusive normal means to κ_{sugar} and is therefore a mere concomitant, and not a demand. Because such contingency plans as $\kappa_{sugar-or-dupe}$, in which the presence of sugar-or-dupe at 108 is a contingency, do not show up as demands for the colony, no tract of behavior by the colony can stand in the intensional acting-like relation to the circumstance *there is sugar-or-dupe-at-108*.

Because $A_{sugar-or-dupe\ at\ 108}$ is not in the colony's remit, conformity to an interpretive scheme involving this plan is not in its remit either, nor is it even an activity it can possibly produce. Because the colony is the only audience in which Brad's dances are adapted to produce interpretive responses, producing such an interpretive response cannot be in Brad's remit and therefore no activity of his, including δ_{108} can be an effort to produce such a response. I will return to the issue of determinacy of content once the account of belief is on the table.

6.6 – Belief

The basic idea about the job-description of beliefs that I develop is that having a belief has the following effect on its bearer: when the believer is trying to get some potential audience to do some κ , a contingency plan for which P is a contingency (and that therefore involves doing something specific in P as opposed to other circumstances), the believer is apt to communicate to

that audience that P. This audience can be the believer herself, as when she reaffirms some important point to herself (aloud or silently) as a reminder.

This idea about beliefs evokes the familiar suggestion that believing P is a matter of having a disposition to say or occurrently think P under certain specified kinds of circumstances. Though it is reminiscent of this behaviorist idea, my account will depart from it because I think the notion of a bare disposition to say or think is too thin. Relying on an agent's bare disposition to say or think P when doing so would be relevantly informative could fail to be the sort of thing we want to call a belief because the disposition fails to reflect the agent's own purposive organization, being rather a product of her circumstances. For example, a God – or even a sufficiently clever neurologist – might devise a way to remotely project into your brain or onto your lips a specific message whenever relevantly informative. In such a case, you'd have a disposition, but its manifestations wouldn't be expressive of your own purposive organization in a way that would incline us to understand you as believing the contents of that message. Such a disposition would instead be an expression not of what you believe, but of what your controller wanted you to act like you believe.

To remedy this issue, I will introduce a “teleologically souped-up” notion of disposition, which I will call *preparedness*.

Preparedness:

Suppose that some functionary o does some π' in an effort to do $[\pi' \gg [Q \mapsto \pi]]$: that is, in an effort to do π in case Q by earlier doing π' . Suppose that having done made such an effort, o has prospects of thereby succeeding, i.e. of thereby doing $[\pi' \gg [Q \mapsto \pi]]$. Then say that o is *prepared to do π if P* , or alternatively, that o *has prepared itself to do π in case of Q* .

Note that to be prepared to do π in case of P does not guarantee that you will do π in case P : only that you have prospects of so doing. Thus, believers might fail to express their beliefs in the sort of circumstances to which the relevant preparedness is keyed for any number of reasons, and

still be believers nonetheless. This is another respect in which this account differs from a straightforward dispositional account: you can still count as a believer in cases where the relevant conditional pattern of behavior doesn't actually hold.

The telltale sign of a believer o of P , I suggest, is that they indicate that P in certain situations. The following is the relevant kind of situation. Suppose some functionary o is trying to get some functionary y – perhaps itself, perhaps someone or something else – to successfully perform some purposive activity. Suppose that the way of successfully doing this particular activity that o is trying to get y to do depends on whether P ; that is, it is some contingency plan κ , involving doing one thing in case of P and other things on other contingencies. And suppose that the relevant contingency P is one that it is within o 's immediate repertoire to communicate to y .

If o believes P , then one might expect o to pursue its efforts to help y do κ by signalling P to y . If o 's signalling effort is successful in that from it ensues conformity by y to the relevant sort of interpretive scheme then y will act like P . And accordingly, y will be driven down the branch of κ corresponding to the contingency P . If this is how o communicates to direct the performance of such κ 's – driving them down the P branch – and if this sort of communicative performance is not a fluke or a result some aberrantly grounded disposition, but a manifestation of a preparedness to behave in just this way, such communications manifest a kind of belief-like state – a take on how things are that manifests in (self)-communicating how things are where practically relevant.

Belief

A functionary o believes P if o is prepared to produce an indicative signal meaning P on the conjunction of two conditions:

- 1) That there is some y (perhaps o itself) such that o is engaged in an effort to get y to perform some κ in $CONT(y, P)$, some contingency plan in which P is a contingency; and

- 2) That some σ that is a signal for y that P , $SIG\langle\sigma(o)|P, u \rangle$ is in the immediate repertoire of o .

Consider, once again, Brad the scout bee's scouting trip. If he encounters nectar he may "get his inner settings configured" in a way that would only be fully capitalized upon if it later caused him to produce δ_{108} upon returning to the hive. Having assumed this condition therefore constitutes Brad's being prepared to perform δ_{108} . Such if done in keeping with normal functioning, would be a signal trying to adapt the contingency plans of the hive that depend on the location of nectar to their "nectar at 108" branches. Because this is the only sort of occasion for a signalling effort to adapt contingency plans relevant to P that can arise within Brad's remit, the preparedness to signal in this circumstance constitutes a preparedness, *tout court*, to signal that there is nectar at 108 in furtherance of efforts dependent on whether there is nectar at 108 – a preparedness to produce that signal "whenever it comes up."

Several points about this notion of belief call for specific clarification. First, it does not require chatterboxhood of an agent nor perfect insight into whether the efforts it is engaged in directing depend on whether P . A functionary that is prepared to ϕ if Q does will not necessarily ϕ if Q without fail; it need only have completed an effort to later ϕ if Q , and to have *prospects* – not a guarantee – of future success in this regard. Thus, an agent may fail to communicate P in a situation where it is trying to get itself or something else to do something in the way suited to P and nonetheless believe P .

Second, it is worth pointing out that "belief" is defined above to include a relatively broad range of belief-like states, and it is only when further conditions are met when the relevant states begin to resemble human belief. For instance, at least two elements are omitted from the definition but are "phenomenologically crucial" to the phenomenon of human belief.

First, Brad the scout bee's belief-like state only prepares him to adapt one very specific sort of contingency plan to the contingency of there being nectar at 108 – it has a chance to make a difference only to what he communicates about where the forager bees are to depart to. A human being's belief in some proposition P can characteristically find expression in efforts to adapt a virtually endless variety of efforts that are contingency plans to the contingency that P . Importantly, this includes the adaptation of contingency plans about what to believe – hence reminding yourself or someone else that P may drive you down the appropriate branch of a contingency plan that says to believe R in case P and $\sim R$ in case Q . Noticing that the characteristic manifestations of human beliefs can be directed at forming further beliefs in this way captures (at least a core element of) the role that belief can play in inference for human beings.

Further, whereas Brad the honeybee's preparedness to signal that there is nectar at 108 can only characteristically manifest in a signal directed at one specific audience whose behavior he is trying to adapt – the colony – human beliefs can characteristically manifest in communications of their content to any number of audience members. And importantly, the audience for our belief expressions can be ourselves – as when we remind ourselves out loud or in thought. The fact that we enjoy beliefs which find expression in self-communication accounts for a dimension of self-awareness that comes along with our beliefs but is not necessarily attached to the dance-producing settings of a scout bee.

Finally, this notion of belief accommodates the platitudinous truth about belief that it is possible to say one thing and believe otherwise. To believe P is to be prepared to indicate P when you are trying to get your audience to *successfully* perform tasks where the correct manner of performance of the task depends on whether P – but not necessarily otherwise. This definition is supposed to be a precisification of the idea that to believe P is to be prepared to say it in those contexts

where speaking the truth about whether P is conducive to what you are trying to do – but not necessarily to say it in other contexts. If you are deceitful, you may be interested in getting your audience to act like P – to perform the branch corresponding to P in one or more contingency plans κ – but to thereby *fail* to successfully perform the contingency plans. Then you may signal that P, without expressing your belief. I believe that the law says that red means stop and green means go. I will speak in accordance with my belief if I am trying to help my audience to drive in a way that doesn't run afoul of the law. If I am trying to get my audience to run afoul of the traffic laws while trying to obey them, my beliefs about the law will not induce me to speak accordingly.

6.7 – Some Further Remarks on the Determinacy of Content

I remarked above that my account of intentional signals – and as a corollary, my account of belief-like states – admits of a good deal of indeterminacy of content. I view this as a feature, and not a bug of the account. This claim might be surprising to someone who takes for granted – as some seem to do – that a theory of content should say once and for all whether, for instance, a bit of brain activity in a frog means fly, or black spot, or frog food (Fodor, 1992, is a paradigm case). Generally, I think that this felt need rests on a mistake that I call the “intensional motte and bailey”. A motte and bailey is a fallacious pattern of argument in which the argument's proponent advances a strong claim but defends only a weaker one with which it is conflated.

Here, the modest claim – the motte – is the idea that our theory of content should make content attributions intensional: intentional state or act ascriptions should not generally admit of intersubstitution of coextensive expressions *salva veritate*. Any theory of content worth its salt should be able to explain how a some mental event can mean HORSE and not HORSE-OR-COW-ON-A-DARK-NIGHT, or why a frog doesn't see a fly as a fly-or-beebee. But it can be easy to slip from this modest idea about the intensionality of intentionality into a much stronger claim. This claim, the

bailey of the pattern of argument, is the idea that our theory of content should make content ascriptions *maximally opaque* in the following sense: it should make it so that whenever a bit of mental (or verbal) activity has the content *x is F*, this same bit of activity cannot have the content *x is G* for *any* G other than F. This is to slip from the idea that content ascriptions should not allow substitution of *whatever coextensive expression we want* into a true content ascription while keeping it true, to the idea that content ascriptions should not allow *any* true content ascription to *ever* remain true if we substitute in *any nonsynonymous coextensive expression*. And this is to pass from the idea that our theory of content should rule out “monsters and misfits” as content clauses – e.g. arbitrary disjunctive contents such as horse-or-cow-on-a-dark-night or fly-or-beebie – to the idea that our theory of content should filter out not just the monsters and misfits, but that it should filter out all the candidate content ascriptions but one.

Once we see that these two ideas are separate, it is easy to see what might intuitively appeal about a theory of content that generally filters out monsters and misfits as unacceptable content clauses without narrowing the range of acceptable content clauses to just one. The sloganeering defense of this idea goes something like this: *you mean to tell me I need to figure out a way of saying whether Brad's dance means there's nectar at 108, or that there's something sweet at 108, or that there's a source of calories at 108, or that there's something worth sending foragers to at 108? Why think Brad is articulate in this way? Why think he is capable with these distinctions? His little world presupposes all of these properties go together. And all the same goes for his audience.* On such a view, the becausal facts of teleological organization rule out monstrous contents like sugar-or-dupe, as discussed above; but where what is in question is a naturally related series of circumstances where Brad's detecting one constitutes his detecting another, we can just say that Brad has no facility with the distinction between these, and that, as a

result his dance is inarticulate as between one or the other meaning. This approach to the determinacy of content is beginning to gain some currency.¹²

6.8 – Whence Determinacy of Content?

The counter-polemic above, accusing polemicists directed at theories of content that harp on indeterminacy of content, of committing a Motte-and-Bailey, overstates the case in a certain respect. While it may be fine to say that representations of bees, or even of frogs, simply possess the kinds of indeterminacies that, polemicists argue, make trouble for teleosemantics, there is a real challenge motivating the polemicists' worry, albeit one that applies more narrowly than they tend to claim. *Our* representations, both in language, and in conscious thought, *are* determinate in the ways that I have suggested we should allow these other representations not to be. We do have an idea of the difference between what is sweet and what is sugary, between flies and frog-food, and even between being triangular and being trilateral. Though the goal of my project is to account for the truth-directedness of belief, and not the determinacy of content, it is worth addressing briefly how my account get deliver the result that human thought and language are as determinate in content as we know them to be, in order to bolster its plausibility.

The ability to distinguish between representations between P and P' , can arise from the audience for which one's signalling abilities have adapted (a reminder, as always, that this may include oneself), to distinguish in behavior, between acting-like P and acting-like P' , to do A_P without doing $A_{P'}$. If your audience can do one without the other, there is room for you to have facility in trying to get them to do so. A sufficient condition for an audience member θ to have this

¹²Neander (2017) expresses the view that there are a variety of ways in which content may be potentially indeterminate, not all of which need to be solved. Recently, in "Should the Teleosemanticist be Afraid of Semantic Indeterminacy?" (2023) Karl Bergman has argued that no, we shouldn't. In large part, he thinks, this is because the kinds of indeterminacy that a teleosemantic theory of content cannot extirpate is "well-behaved" in something like the way I have just described.

ability is that they can perform every plan in $CONT(o, P)$ that is demanded of them by performing the corresponding α_P without performing any $\alpha_{P'}$, the branch corresponding to P' , in any of the contingency plans in $CONT(o, P')$: taking P -branches of contingency plans in which P is a contingency without at the same time taking corresponding P' -branches.

Forager bees who respond to Brad's dance cannot do this. The behaviors that the presence of nectar occasions are just those that the presence of usable sugar occasions – and if they are doing one, they are doing the other. On the other hand, we, who make many contingency plans of all sorts may have different specific plans for cases that typically go together but may come apart.

Historically, we have sought vitamins by seeking food. Now we have at our disposal usable vitamins without food, in the form of tablets. We have built laboratory equipment that is sensitive to the presence of vitamins, even when it comes apart from the presence of food, and have developed techniques to make vitamins without food and we exploit both of these to fill labeled bottles full of vitamins without any food; they can then, because of their labeling and conventional shape, be treated differently than food without the use of special instruments. The people at the factory treat these differently than food, at the drug store they're treated differently than food, and when they're in your medicine cabinet at home you treat them differently than you would treat food in your medicine cabinet.

The explanation of how we can form representations that are determinate as between representations about vitamins and representations about food means depends on our ability to discover means of making ourselves differentially sensitive as between the two. This approach obviously does not work to explain how our representations may be determinate as between contents that necessarily covary – for example, of how we can think determinately about triangularity as opposed to trilaterality. Here, the hyperintensionality of the “because” that figures in

the notion of purposive organization I have been employing, comes to the rescue. Because we are capable of strategies of differential responses – we can treat something as a triangle by noting angles and counting number, or by noticing edges and counting their number – we are capable, by employing one of these methods or the other, of treating something in a particular way *because* it is triangular or of treating something a specific way *because* it is trilateral, without necessarily doing the other. Likewise, we can organize ourselves around contingency plans that dictate that we treat shapes differently depending on their number of angles, without inducing any purposive organization around differential responsiveness to number of sides, and vice-versa. And because we design ways of communicating to trigger either of these responses or the other, we can send signals that indicate triangularity as opposed to trilaterality, or vice-versa.

Notice, further, that being able to harbor beliefs, as opposed to merely transmitting signals and immediately (forgetting) can provide the opportunity for extensive further determination of content. When a signal leaves its sender, it is an effort to make use of whichever audience interpretive schemes it is an effort to make use of, and that is that. A belief, however, might issue in many signals, which take place over the course of a process of development of more sophisticated, more finely determining interpretive schemes. A belief's directedness not just at making use of such interpretive schemes as there are, but at potentially making us of such as might develop, allows it to possibly exhibit a kind of determinacy that outruns our present ability to make distinctions.

§7 – What Comes Next: Truth Normativity and Universal Planning

Above, I have elaborated proposed notions of the kind of conditional requirement involved in a hypothetical imperative as well as a notion of belief. Putting these together, we get the following formulation of the Truth-Normativity thesis about belief of some proposition P.

The thesis asserts, first, that there is a conflict-sensitive way of making a signalling effort instructing you to believe P and not to believe P if not- P . Elaborating this requirement based on my notion of belief from above, this would amount to an effort to make it so that you are prepared to signal P in the relevant circumstances if P , and that you embody no such preparedness otherwise.

Second, the thesis requires that this effort have prospects of succeeding in a conflict-sensitive way, i.e. of making it so that you have the relevant state of preparedness without thereby lessening your prospects for other relevant aims. Finally, it requires that this achievement have a potential point; that is that there be no level at which, when we chase outward from means to ends, no matter how far we go there is some use to which this communicative preparedness has prospects of being put in the course of normal functioning.

I want to point out that any suspicion that this condition is not met for believing some P iff P , that the aim of doing so is conflicted (no conflict-sensitive variant), that it is precluded (no prospects), or that it is conflicted (no point), seems that it would have to derive from limitations on our repertoire. For the issues conflict and for that of prospects, this point is trivial. Limitations on our ability to attain the relevant state of belief, or limitations on our ability to do so without conflicting with our other goals (or limitations on our ability to try for this by signalling) would be precisely a matter of lacking means at our disposal to achieve the result, or to do it in a way that does not diminish our other prospects. For the of pointfulness, notice that beliefs are, as such, directed at executing contingency plans for which there is a point of doing them. Thus, as long as there is some plan κ (that itself has a point) with a contingency P for you to adapt, there is a point in your believing P . Thus, we can only place limits on the range of cases in which believing truly has a point if we can place limits on the range of contingency plans we might adopt as aims to which a true belief that P might be relevant.

It is just this kind of limitation on our repertoire that I want to deny holds in any concrete or substantive way. Sure, there are things that we cannot do because it is physically impossible that we do them at all, in the course of our normal functioning or not; perhaps travelling at faster than light speed is one such thing. But we cannot, I will suggest, place any limits on what it might be within our power to accomplish that goes any farther than the limits on what it is possible may merely happen to us. The *Universality Thesis* says specifically for us that we do not know of any plan ρ such that it is possible that we do ρ but doing ρ is not in our repertoire – that is, that it is possible for the doing of ρ to be something that merely happens to us but not something we purposively accomplish.

In the concluding chapter to follow, I will argue for this principle expressing the unboundedness of our potential, deriving it from the connection of our individual purposive organization to our collective purposive organization, which makes it so that the extent of our individual repertoires reaches as far as our collective repertoire does. I will go on to explain how this absence of substantive boundary, along with the presence of a “cost-free” way of aiming at truth by signalling (the kind that is made explicit in the dialogue between Helen and the Hare from chapter 1), renders Truth-Normativity a correct expression of a hypothetical imperative governing our form of life: best efforts at leading a human life entail efforts at believing as things are, in as unrestricted a sense as possible

Chapter 4 – Universality of Repertoire and Truth-Normativity

In the previous chapter, I suggested that we can establish that we enjoy beliefs subject to truth-normativity by establishing the Universality Thesis: that there is nothing that might happen to us that we can rule out of our repertoire, the set of things we might be able to do in the course of our normal functioning. The main goals of this chapter is, first, to argue that the Universality Thesis is true, second, to explain in a little more detail how the universality thesis vindicates truth-normativity. I conclude by briefly tying up some loose threads.

§1 – The Argument for Repertorial Universality

The Universality Thesis is a thesis about our repertoires: the range of things we may do in the course of our *normal* functioning in response to the challenges we face. The notion of a repertoire, and UT in particular, draws on the distinction between what we might “do” in the extended sense where accidentally tripping and falling is something you do, and what we do *on purpose* in a broad sense of “on purpose, in a broad sense of “on purpose,” meaning it is not by accident, it is not something that merely befalls them. An element ρ of σ 's repertoire need not be doable by σ on purpose in the sense that σ can do it *consciously*, or *deliberately*, or *by first deciding to*: can openers and bacteria have repertoires, too. A repertoire contains those activities that we might do on purpose in this extended sense.

The Universality Thesis claims that we cannot draw substantive distinctions between the kinds of things that might merely happen to us and the kinds of things we might do on purpose (in the extended sense of “on purpose” in which we grew our teeth on purpose).

The argument I will make for it centers around the oft-noted ideas that human technological and social development is reflexive and cumulative: a previous generations cultural advances typically persist into the next, and can be built on, and further, can themselves enhance the available

methods for building on the previous advances. What I add to these platitudes is a way of thinking about how this cumulativeness and reflexivity is not just something that happens in human culture, but something around which human culture is organized, an operative purpose thereof. More specifically, I understand this purposiveness as grounded in our ability to bring our purposive activity under representational control, which confers upon our form of life a *self-developing* character.

The course of the argument runs as follows: I outline a paradigmatic way in which it can be argued that some ρ is *in* an σ 's repertoire (1.1-3). I note the options for the claims that would have to be defended in order to block such an argument (1.4). I make some observations about how the use of representational activity in controlling purposive activity can enhance the repertoire for performance of that activity (1.5). These enhancements create an impediment to the blocking move which would exclude a plan from a repertoire. Because of the respect in which the devising and use of means of symbolic control is general, the *Collective Universality Thesis* – the counterpart of the universality thesis that applies to us collectively obtains (1.6-7). And because each of us individually is organized around collaborating in plans that are operative for groups of which we are part, I will argue, we can get from the collective universality thesis to the individual universality thesis (1.8).

1.1 – The Paradigm for Repertoire Inclusion: The Repeatable Precedents Rubric

An element in σ 's repertoire ρ is something it might perform normally. The intuitive gloss of what it is to perform normally is for σ to do ρ in the way that σ evolved, learned, or was designed to. In such cases, there is a nexus of explanatory connection between the events by which the marks of purposive organization – the features of the beneficiary in virtue of which the relevant function is operative – became established as such, and a possible future performance. I'll focus on a simple rubric for connecting such adaptation-entrenching events and future performances and argue that it is more flexible and pervasive than it initially appears. My argument will not require the claim that

the fact pattern described by this rubric is the only activity ρ may end up in a repertoire. The discussion may serve to show, however, that this suggestion is not implausible.

The simple fact pattern in question involves two parts. First, that there be historical performances of a relevantly related function, explaining why the marks of purposive organization are what they are. I will call these performances *entrenching performances*. And second, there are possible future circumstances in which the explanation of how in the past cases, these marks led to or ensured successful performance of the function would also work to explain how they lead to or ensured the performance of the function again, and in particular, would lead to the performance of the function by means of σ 's doing ρ .

Here is the first part of the fact pattern in a little more detail: suppose that there is some operative plan π for some U of which σ is a part. In the past, U , or its historical counterparts had some properties F which were more conducive to doing π than some alternatives, F^* , and because of this comparative conduciveness, in past cases, being F led to successful performances of π more reliably than being F^* did, which in turn led U 's being F , rather than F^* . Call these past cases of successful performance of π that lived up to F 's comparative conduciveness and explain U 's F -ness today *entrenching performances*.

Now the second part of the fact pattern: suppose that it is possible for there to arise circumstances that are similar to the circumstances in which the entrenching performances occurred. Suppose, specifically, that circumstances could be similar such that U could do π and the explanation of *how* the marks of purposive explanation entrenched in the entrenching performances led to doing π , would also explain how they led to U 's doing π in this case. And suppose, further, that in these circumstances, U 's doing π would require that it do so by means of σ 's doing ρ . Such a

possibility is the possibility of o 's doing ρ normally, and so the fact of such a possibility places ρ in the repertoire of o (more specifically, in its repertoire for U 's doing π).

Call this fact pattern the *Repeatable Precedents Rubric*. For products of direct natural selection, or behaviors learned through trial and error, the way in which this rubric applies is simple. There is no distinction between π and ρ or between U and o . We can establish that pumping blood is in a heart h 's repertoire simply by looking at past circumstances in which ancestral hearts pumped blood, and finding that the way in which they managed to do so is like the way in which h 's marks of purposive organization would lead to its doing so. If a bear learns a specific motor sequence that opens a trash can we can cite the possibility of future performances of this motor sequence that work just the same as the entrenching past performances to establish that this motor sequence is in her repertoire.

1.2 – The Repeatable Precedents Rubric and Novel Performances

The fit is slightly more complicated for potential *novel* performances ρ in the repertoire of o . The Wright brothers had heavier-than-air flight in their repertoire before they made their first plane, but the pattern of facts that make this the case does not involve mama and papa Wright designing flying machines. If Brad the honeybee is the first of his kind to encounter nectar at a particular orientation, performance of the corresponding dance is still in his repertoire, despite the lack of any prior performances of just that dance. Nonetheless, there is an attestation to the membership of these activities in the functionary's repertoire that fits the Repeatable Precedents Rubric.

In these cases, unlike in the simple cases above, the precedented operative plan π is distinct from the target repertoire member ρ . The precedent in question is a *contingency* plan of which ρ is a specific branch. In Brad's case, this works simply: Brad's ancestors evolved to execute the *contingency*

plan $\pi = [P_x \mapsto \delta_x]$. And circumstances Brad finds himself in may well be just like the circumstances of the entrenching performances of π - those in which his ancestors did π and in virtue of which he is now organized around doing π - so that he has an opportunity to perform π normally, drawing on the same elements of purposive organization to do his dance as his ancestors, with those elements securing correct performance for the same reasons that they did for his ancestors. This is the first time, however, that circumstances are such these mechanisms functioning in the usual way would secure the performance of π in a specific determinate way, namely by performing the specific branch - say, δ_{108} - corresponding to the location at which nectar was encountered.

The story of how flight was in the repertoire of the Wright Brothers fits the Repeatable Precedents rubric in a yet subtler way. Here, the key fact is that the Wright Brothers achieved heavier-than-air flight by learning principles and techniques of engineering. Someone who has learned such principles is thereby organized around an operative contingency plan where the contingencies are circumstances in which some engineering problem comes up and the corresponding branches involve designing or producing items that answer to these problems. Achieving heavier-than-air flight by building a machine capable thereof is thus a branch of this contingency plan.

However, merely assimilating heavier-than-air flight to this contingency plan does not show that it is in the Wright Brothers' repertoire in accordance with the Repeatable Precedents Rubric. The relevant engineering knowledge in virtue of which the aforementioned contingency plan was operative for the Wright Brothers was not plausibly entrenched in their minds by a process of trial and error, either by the biological evolution of the cognitive basis of these principles, or by their learning them through trial and error from scratch. Rather, they learned them, in part, by receiving

the right sort of intelligent instruction. Thus, the contingency plan whereby they are to employ such-and-such principles and techniques of engineering when they want or need or decide to make something is *itself* a branch of a contingency plan. This higher-level contingency plan around which the Brothers are organized dictates that they respond to various kinds of instruction in correspondingly appropriate ways. One contingency involved in this plan is that they find themselves receiving instruction in the engineering techniques in question, with the corresponding branch being that they acquire competence with these principles.

Even this learning contingency plan that the Wright Brothers employed to learn how to engineer and, in turn, to fly in a heavier-than-air machine, is not plausibly directly evolved; a person is not born with needs first to acquire the tools needed to respond positively to the relevant kind of instruction in order to learn how to be an engineer (for instance, an adequately expressive natural language, adequately expressive mathematical language, competency in the behavioral practices of a good pupil, and so on). So the operative plan of exhibiting the kind of pupil-hood that would allow them to learn engineering is itself a branch of some meta-learning contingency plan.

This point may even iterate several more times. But eventually, we should come to a level of description at which the tools the Wright brothers draw on are their directly biologically evolved cognitive capacities. Heavier-than-air flight is thus a branch of a contingency plan at the bottom of a series of nested contingency plans, where the highest-level contingency plan is operative for the Wright Brothers in virtue of their biological evolutionary history. Their cognitive equipment, under the right specific circumstances, works just the way that their ancestors' cognitive equipment did in entrenching performances, with the result, in these circumstances, that the brothers are prepared for the first grade. If things continue to function in keeping with this particular purpose, then, plugging

in further environmental parameters, they will eventually become prepared for high school, and then prepared to tackle engineering problems, and eventually, to fly.

Millikan expresses something like this idea about intelligent behavior and its products nicely in *LTOBC* when she writes (47-8 emphasis and capitalization in original):

...humans not only learn, they also learn how to learn (e.g., develop concepts) and possibly even learn how to learn how to learn (e.g., develop methods of concept formation). But however flexible the human nervous system is, containing systems that are instructed or programmed by still other systems, still there must come an end to flexibility. Both the outermost systems and principles involved and the kinds of flexibility possible in programming more inner systems must be inherent in the basic brain – the original product of evolutionary design. There must be a way of describing the inherited nervous system and certain very general environmental conditions (e.g., being in the context of a human “society,” given an appropriate definition of “society”) under which the nervous system has historically operated such that a general explanation can be given of *how* that system has normally accomplished those ultimate functions that have led to its proliferation.

Although I have just noted a nice echo in Millikan of the kind of story I have rehearsed above, there are two respects in which the idea expressed she expresses in the just-quoted passage fits awkwardly with my purposes. The first arises in Millikan’s talk of how we are “programmed.” The term suggests a kind of rigid determinism wherein, at least *ceteris paribus*, relevantly like inputs lead to relevantly like outputs, without notable variation. Though the running example contingency plan, a bee dance, may fit this description, this notion of a program does not describe all contingency plans, as I am understanding them. As I understand them, contingency plans can be quite “mighty:” functionaries operating in keeping with them need not respond to a specific contingency by performing the associated branch with perfect, or even particularly high reliability. When a paintball player aims and fires, she is operating in keeping with a contingency plan to send a paintball to location x when her opponent is at location x , even if it typically takes her a number of shots to hit her opponent and so execute this plan. Talk of a “program” of the form *if opponent is at location x then send paintball to x* suggests a system that satisfies this condition virtually always, whereas

she only does so once every five shots, say. Further, a functionary organized around a contingency plan need not always perform a given branch in the same way but may be capable of responding to a circumstance P by performing the corresponding branch π in a heterogeneous range of ways. For instance, a designer might sketch out several ideas for a dress in response to her client's instructions. In doing so, she may in each case be pursuing the same branch of a contingency plan that pairs instructions with dress designs but satisfying it very differently.

Second, relatedly, and even more importantly, Millikan's talk of an "end to flexibility" and her framing in terms of a process of description that starts with the wide variety of human behaviors and traces back to a level at which things are uniform looks suited to exactly the opposite of the result I want to aim at, where it looks like we ground out in something rigid and closed-ended. This difficulty is equally present in the way I made my way through the phenomena in discussing why flight is in the individual repertoires of the Wright Brothers. I don't take back what I said there, but this way of working through the territory is not especially conducive to indicating how the Universality Thesis might be true. To do that, we need a way of looking at things that does not begin with the varied phenomena working backwards to the level of uniformity. We instead need a perspective that begins with a level at which entrenching performances are uniform, and works outwards from these uniform precedential performances, to yet more and more varied performances for which they are precedents, and which does so in a way that indicates that going in *this* direction, working from uniformity forward to variation, we can go on getting more varied as far as we like. To achieve this perspective, and in particular, to get the idea that steps of further variation can go on as far as we would like them to, we need to move to the perspective of our collective purposive organization.

1.3 – The Repeatable Precedents Rubric and Novelty in Collective Repertoires

It is one thing to say that flight is in the repertoire of the Wright Brothers. It is another question whether heavier-than-air flight was part of the collective human repertoire, something available to us to tackle the challenges we'd characteristically face, before the Wright Brothers, or before the Industrial Revolution, or before large agricultural societies with variegated social structures.

And likewise, it is another question, if flight was in our collective repertoire at any of these points, whether it was in our collective repertoire because of a fact pattern that fits the Repeatable Precedents Rubric. I have suggested above a way of thinking about how, if the Wright Brothers' brains do the usual human brain thing, in conditions generically like those which human brains are adapted to work in – though specifically inflected by the idiosyncrasies of their circumstances – can eventually fly in something heavier than air. The relevant circumstances that allowed the Wright Brothers to do the trick, though requiring what was at the time of the Wright Brothers nothing out of the ordinary beyond the good luck of finding the right idea, involved all sorts of features of accrued technological progress that would have been unthinkable to paleolithic humans. So it might seem difficult to say that the path from arrowheads to aeroplanes can be described as something that happened in the usual course of business, rather than its being a sheer accident that a species exhibiting the kinds of abilities our species did thousands of years earlier should achieve flight.

Nonetheless, I do think that flight has been in our collective repertoire for a very long time, and I think that the fact pattern that made this so falls in line with the Repeatable Precedents Rubric.

First, notice that on the theory of purposive organization I advance two chapters ago entails that populations as a whole are purposively organized. Natural selection makes populations one way rather than another – composed of individuals with this phenotype rather than that phenotype –

because the former way is conducive to certain kinds of outcomes (in particular, that its members function in a certain way).¹

A strategy that was entrenched in our species long ago was to be made up of individuals capable of making tools, of copying one another, and of cooperating in coordinated behaviors with variegated roles – that is in which one of us does one thing and another of us does something else – and importantly, armed with natural languages, which we are interested to use in supplementing pulling these feats off, as well as to remind one another of what we do and how we do it, lest we forget. This is a very general ability of the species, and the kind of activity that the marks of purposive organization tailored to these abilities conduce to can take any number of forms, depending on the kinds of achievements and practices already in play in the moment it is operating. Though it produces new outcomes as new developments accrue, when described abstractly as above, it is in one way just a matter of turning the same crank over and over again.

Though this line of thought so far has only been an argument that flight is in our collective repertoire, not that everything is in our collective repertoire, I hope that it is suggestive of the arc that the overall argument will take.

1.4 – How to Rule a Plan Out of a Repertoire

As I've argued in previous sections, one kind of fact pattern that has a wide-reaching ability to place a plan in a functionary's repertoire is the Repeatable Precedents Rubric. Thus, to rule a plan out of a repertoire, we need, at least, to rule out that a fact pattern in keeping with the Repeatable Precedents Rubric places it in a repertoire.

¹ Indeed, the population level is the most basic level at which CCET functions apply. Natural selection makes the polar bear population have the trait distribution of being uniformly white rather than, say, uniformly brown, because of conduciveness to camouflage. In turn, this explains why the population spits out new individuals that are white, rather than brown.

One way to argue for this is to argue that there is no plan π that is operative for θ that could be accomplished by doing the ρ in question. For instance, it is not in the repertoire of a fish to travel to the center of the Sun. And to see this, we need not even go as far as thinking about normal performances. We can simply consider the range of plans that might be operative, might be purposes around which a fish is organized. And then we can ask whether any of these could, under any circumstances in which a fish might find itself, be such that they would be furthered by traveling to the center of the Sun. The answer is, probably, not.

A second way to block the Repeatable Precedents Rubric's path to establish an element of a repertoire is to argue that even if there is some π that is operative for θ and π could be done by doing ρ , that the marks of purposive organization that are associated with π

- 1) could not cause or ensure $\rho(\theta)$ at all, or
- 2) couldn't cause or ensure $\rho(\theta)$ in conditions where doing ρ would be a means to doing π and could only do so in other situations, or,
- 3) could cause or ensure $\rho(\theta)$ in a situation in which it would be a means to π , but the reasons they conduce to doing ρ are not of a piece with the reasons that these marks of purposive organization conducted to doing π in the entrenching performances, so it is sheer coincidence that they caused the relevant result.

For example, an effort to rule predation of an eagle in flight out of the repertoire of a bear might proceed according to option 1) above. An eagle soaring above a bear might make a good meal for her, it is not in her repertoire for her to catch and eat it, because in any version of the case that is not extremely aberrant (e.g. she is shot out of a cannon) it is not possible for her tools of predation – her teeth and claws – to reach the sky.

It is hard to come up with cases that fit 2) well, but it is worth noting that there is logical space for it.

Move 3) could be made to block the Repeatable Precedents Rubric as part of ruling out that being collected and kept as a pet by humans is part of the repertoire of a newly discovered brightly colored species of snake. It might be that the snake is brightly colored as a way of deterring predation, and being captured and kept as a pet would be a way to avoid predation. If such a snake is kept as a pet because of its bright markings, then its adaptations for avoiding predation cause it to be kept as a pet, and thereby to avoid predation. But the markings don't contribute to its getting captured in the way they contributed to avoiding predation in the evolutionarily significant cases. So the entrenching cases of deterring predation through colorfulness do not set a precedent that being captured due to colorfulness repeats. Thus, being kept as a pet by humans can likely be ruled out of such a snake's repertoire.

1.5 – Repertoire Enrichment Through Symbolic Control

I have just given some indication of what is required to rule a plan ρ from being included a functionary's repertoire in keeping with the Repeatable Precedents Rubric. In this section I will discuss the characteristic effect that being able to use indicative and intentional signals has on a functionary's repertoire. In the next section, I discuss the way in the availability of such a method to a functionary complicates attempts at repertoire-exclusion.

My idea about the use of representational devices to enhance purposive activity is broadly inspired by Sellars (1954, 1981). On his view, as on mine, there is a level at which activity may be purposively organized, and hence follow a certain pattern on purpose, though it is not controlled by representations that it is to proceed according to this pattern. Sellars calls this sort of activity "pattern-governed behavior" (1954). Representations of a pattern of behavior that can induce that pattern of behavior have the effect of providing a supplementary pathway to the pattern of behavior in question, acting as a backstop or expanding its reach. The deployment of such representations as

a means to backstop the relevant pattern of behavior is itself pattern-governed behavior when it occurs. Thus, it may also be subject to backstopping by further, higher-order control. One way of thinking about the essence of the view I'm advancing here – and Sellars hints that he shares this idea – is that, at the level of potentiality, this hierarchy of representation-mongering control goes “all the way up.”

Consider the game of chess. In most positions a player has available to her one or more dozens of legal moves. In choosing an effective move, she will want to know how her opponent could most effectively respond. And again, this may require a choice from among one or more dozens of legal moves. Which response from her opponent would be most effective in turn depends on what our player would play on her *subsequent* move, which again introduces potentially dozens of choices. It very quickly becomes impossible to work through every possibility a given number of moves out. So a player must decide which moves to evaluate and which moves to ignore. An experienced chess player can do this “by feel;” having trained herself through experience, certain moves will “stand out” to her as worth checking. This is a purposively organized process: an experienced player's brain is disposed to notice more salient moves in virtue of having practiced chess. But this is not mediated by conscious thought: the first thing you are aware of in such a process is the move or moves that it occurs to you to evaluate. Generally, the stronger the player, the better they can perform this identification by feel. For less advanced players, it is often useful to interpose explicit reminders that employs linguistic representations of different kinds of chess moves. Beginners are often reminded to look for “checks, captures, and attacks:” moves that will threaten the opponent's king, take one of the opponents' pieces, or place one of their pieces in a position to take an opponent's piece on the next move. In evaluating their opponents' likely next moves, they are also asked to look for the checks, captures, and attacks their opponent can play against them. This introduction of a control based on the use of classification mediated by symbols

may improve a player's prospects of identifying the right moves to analyze. Relying on the purposive organization generated just by the beginners' (limited) experience with the game may be useful to some extent, but having a backup method can often improve the reliability with which they succeed.

These sorts of repertoire-enhancing strategies can have quite dramatic effects. The use of representations may allow the substitution of highly precise strategies for imprecise ones. A dog who has swallowed a foreign object and has an intestinal obstruction might have some physiological adaptations which are organized around clearing it, and which have some chance of working. We, on the other hand, who are interested in helping our dog feel better can find a veterinarian who can identify the problem exactly and intervene by surgery.

Not only can the use of symbols improve prospects in purposively organized activity where there already are prospects; it can also introduce new methods of achieving old goals, and thereby create prospects where there otherwise would be none. Humans, like other animals, are organized around behaviors and physiological responses that regulate their temperature, maintaining it within an appropriate range. One element of our repertoire for achieving this regulation, one which we share with many other animals, is behavioral regulation based on perceived body temperature. When we feel warm, we tend to seek out shade, cool water, and the like, and when we feel cold, we tend to seek out heat sources, and so on.

This behavioral toolkit serves us well for thermoregulation in most circumstances. But paradoxically, it is not uncommon to find people who have removed all of their clothes shortly before freezing to death; at a certain stage of hypothermia, people begin to perceive (mistakenly) that they are too hot, and may behave accordingly. In such a circumstance, the usual repertorial strategy of trying to warm up when you feel cold and cool off when you feel hot is no longer a means to successful thermoregulation. So if this is the only strategy in your repertoire – or even your

only strategy beyond the unintentional physiological responses to change in temperature – you do not have prospects of thermoregulating in these circumstances. However, matters are different if you have in your strategic toolkit not just feelings-based behavioral strategies, but also strategies that are informed by *having heard of hypothermia* and *being able to estimate numerical temperatures*. In such a circumstance, you might be able, even while feeling hot in the throes of hypothermia, to hold the belief that you are hypothermic, or that the temperature is very cold, as well as the belief that the way to keep your temperature within a safe range is to try to warm up. If you can then remind yourself of these things, you can act contrary to and in spite of your feeling of overwarmth and may be able to survive. Notice that here the use of representational capacities not only generates a behavior which would otherwise not be available to you in the course of normal functioning, but also a *state* you wouldn't otherwise have achieved: being warm.

Notice, further, that the augmentations mentioned above not only provide new tools in our repertoire, but also occasion new demands. For example, because we are capable of representing the numerical temperature and forming contingency plans accordingly, the fact that there are situations in which sensitivity of the numerical temperature would meet our needs in turn creates an occasion to make ourselves sensitive to the numerical temperature, i.e. to make thermometers.

I have made some specific points, appealing to examples, about ways in which the use of representations augments a repertoire with new means to ends, thereby improving prospects and creating prospects where there otherwise would be none, as well as by creating new demands to go along with the related new abilities. I want to conclude this section with some more general remarks about how representation works as a repertoire-expanding tool.

In the *Philosophical Investigations* (1953/2009) Wittgenstein makes an analogy between elements of language (specifically individual words, but we can extract lessons about sentences) and handles in

the cabin of a locomotive. He is following up on the point that from the previous paragraph that, while at some level of abstraction, we can call our various words “tools,” to stop here would be to paper over the heterogeneity of the ways in which words help us to interact with the world. He writes (§ 12):

It is like looking into the cabin of a locomotive. There are handles there, all looking more or less alike (this stands to reason, since they are all supposed to be handled.) But one is the handle of a crank which can be moved continuously (it regulates the opening of a valve); another is the handle of a switch, which has only two operative positions: it is either off or on; a third is the handle of a brake-lever, the harder one pulls on it, the harder the braking; a fourth, the handle of a pump: it has effect only so long as it is moved to and fro.

There is a clear negative contention here about language: the outward uniformity of words – or more pertinently for our purposes, sentences – disguises the heterogeneity of what good they do us and how. But the analogy also makes a positive point: it is no accident, Wittgenstein points out, that the handles all look the same, though they do very different things. This uniformity allows the wide variety of functions he mentions to be performed by way of interface with a single appendage, the human hand.

This idea about the benefit of a uniformizing interface can be extended to *interactions* between pieces of language, and in particular, to sentential expressions. We can alter the analogy and speak instead, if we like, of components of a complex LEGO™ machine with various moving parts. Various as the parts of the machine are in their purpose and manner of functioning within the overall machine, they have in common that they all may be stuck together by uniform little nubs. Likewise, as different in various respects as the work that sentential expressions do is, they may nonetheless be fitted together in complex sentential expressions or arguments. Appreciating the importance of this positive fact about language is what motivates the urgency of solving the Frege-Geach problem for expressivist accounts of moral claims, for instance: we might think that there are

deep differences between the ways in which a moral assertion and an empirical assertion figure in our interactions with the world, but to do justice to the use we make of these kinds of expressions *together*, we must say something about how they can be combined in conditionals and put together in lines of reasoning.

For our purposes here, this point gives rise to the profound *unifying* and *productive* effect on a creature's purposive organization of a being able to do something like wield a representational system. Early in the previous chapter, I drew our attention to the importance of the fact that an evolved creature doesn't simply hold all of its purposes in an unstructured pile. But at the same time, an evolved being gets its various purposes from disparate bouts of selection that took place in response to disparate pressures at different points in its evolutionary history. These may not all be optimally coordinated and organized. A style of functioning that might be useful in various circumstances an organism might face may be available in only one of those circumstances per its evolutionary endowment. Or a style of functioning suited to one kind of challenge may interfere in an unnecessary way with one suited to a separate challenge, when these challenges overlap, nature not having yet stumbled upon a way of getting the strategies for meeting them to harmonize.

An animal equipped with the capacity to control its interaction with the environment by wielding representations of this environment's features, and of the behaviors it can produce – of exercising the kind of “retreat to representational control” I discussed above, the occurrence of a representation of what to do or how things are may be deployed to bring functioning correspondingly into line – can help to improve upon this predicament. If you have a way of making arbitrary structured combinations symbols for circumstances or behaviors, which (are meant to) correspond to – be produced by, and produce, respectively – structured combinations of those circumstances or behaviors you can represent, then you have available to you a way of making new

complex behaviors, and new complex kinds of circumstance-responsiveness by simply putting together your representations accordingly. And if you have facility with correlations between features of those representations and noteworthy features of the relevant behaviors and environment, you may be able to do better than mere trial and error in generating new behaviors by generating new representations thereof – you can generate new behaviors by performing calculations.

This idea should not be read to convey chauvinism about human beings or about public languages. The activity that goes on in the brains of non-human animals – regardless of whether its fundamental workings are most perspicuously described as operations in a “Language of Thought” – clearly accomplish these productive and unifying tasks. And it is also not enough to establish the Universality Thesis. For all that this point shows, there may still remain the unwelcome element of the line of thought about intelligent behavior I quoted at length from Millikan above, that at some point flexibility comes to an end, when a creature reaches the limits of what “hard-coded” representational systems can do. The missing ingredient, to foreshadow, is an open-ended ability to devise new schemes of representation whose repertoire-expanding power goes beyond those of our old ones. But all the same, these considerations should indicate how representing expands our repertoires in the direction of universality.

1.6 - How Representational Control Detains Attempts to Exclude Plans from Repertoires

Above, I’ve tried to indicate how representational capacities can expand a functionary’s repertoire to contain plans that it would not otherwise contain. In brief, if a creature possesses an ability to represent the contingencies of its environment and structured plans in which ones behaviors could inhere, a creature may be capable of executing plans by coming up with formulas corresponding to them – including executing contingency plans by coming up with formulas

corresponding to the contingencies. In this section I suggest that this consequence of a functionary's representational capacities poses a complication for any attempt to rule a plan out of its repertoire.

The basic idea is that introducing considerations of representational control causes a retreat to a more abstract level at which the question of whether there are precedents which the doing of ρ could repeat can be asked. Our focus shifts from some underlying level of purposive activity and what the organization directly around that activity could generate in keeping with past precedents, to the question of higher-order activity of controlling and generating new strategies for that underlying purposive activity, and whether something like ρ could be spit out in keeping with this process.

In order to rule out the inclusion of ρ in some σ 's repertoire in keeping with the Repeatable Precedents Rubric, we need to make a case that, of all the operative plans π that prevail upon σ , and of all the entrenching performances that made for the marks of purposive organization that ground the operativity of those π , there is no nexus connecting the ways in which those marks led to or allowed the performance of π and a way in which those same marks could cause or ensure the performance of ρ , as a means to performing the same function.

When some purposive activity, π is subject to symbolic control, this generates more options ways in which a performance of ρ could count as "more of the same." The fact that the way in π was done in the past entrenching performances has nothing whatsoever to do with ρ doesn't settle the matter. σ may have evolved to do π in one way, but at the same time have *become organized around the use of symbols to find new ways of accomplishing plans* of some kind of which π is an instance. This shifts our attention from the entrenchment that is responsible for the organization around π to the entrenchment that is responsible for the organization around the method of symbolic control that can apply to such plans as π . If the means at σ 's disposal *to make new plans* as means to accomplish

extant goals by generating representations of these new plans could operate in the usual way and generate a case of doing $\rho \gg \pi$, then ρ is in σ 's repertoire. It is in keeping with something like this explanation that engaging in warmth-seeking behaviors to avoid dying of hypothermia, even when you feel hot, is part of your repertoire. It is something you could do, not by following your usual somatic-behavioral patterns, but by operating in keeping with your usual methods of *instruction-following*.

For much the same reasons, the question of whether a demand for which ρ might be a means is in σ 's remit undergoes a widening of focus when we consider methods of representational control. In particular, we can ask not just what needs can come up in operating our underlying capacities as usual, but also about what needs might come up as a result of operating the means of representation-mongering control as usual, e.g. the need for thermometers as a result of the ability to representations of numerical temperature and to make contingency plans on which they depend.

1.7 - The Collective Universality Thesis

I have claimed above that when representational control is at issue, we can shift the focus of the question of whether past entrenching performances set a precedent for some putative repertoire member ρ from an underlying level of purposive activity to the higher-order activity of generating behavior to fulfill antecedently existing purposes.

The idea behind the Collective Universality Thesis is that we are organized around taking merely purposively organized behavior not controlled by representation and making new methods of representational control to apply to it. We make new modalities of language, new scripts, new mnemonics and calculuses and word-games, new theoretical frameworks and new formal methods, new vocabulary, and so on. The more we can represent, the more we can do by means of representation. And we are organized around taking these methods of representational control and

applying them to control activities which we previously without such representational mediation – for instance, in the predicament of hypothermia from above.

And this capacity is targeted at our purposive activity quite generally. This means that it is targeted at our representational activity itself, making this capacity *self-developing*. This means that when we ask whether business as usual could result in the purposive doing of ρ , for us, we need to look not just at what our present methods of representational control – the kinds of techniques and the kinds of underlying activities to which it can be applied – might spit out. The “step-back” applies iteratively: we need to look at the ways in which we might develop our methods of representational control, and what innovations these methods *thus* developed might spit out. But then we need to ask about the ways in which we might develop our methods of developing our methods of representational control, and so on. Our self-developing capacity to take underlying purposively organized activity and bring it under representational control – which applies to the methods by which it is practiced itself – makes us indefinitely capable of deflecting away from bottoming out at any level of abstraction where a novelty would have to be a genuine accidental novelty, rather than a newly developed variation on the same theme.

This rather abstract and theoretical argument is bolstered by the positive considerations glossed above in connection with the idea that flight has been in our repertoire for a very long time before we could do it. The way we live, and the kinds of organized activity we are capable of have changed radically throughout human history, and are changing faster and faster all the time, as technological progress accumulates. But fundamentally, this can be understood as a matter of the application and reapplication of one strategy: taking what we do now and finding new ways of thinking about it in order to find new ways to do it better.

1.8 – Individual Universality from Collective Universality

The idea from above is that we can keep turning the crank of taking our predicament, and the strategies we follow without representational mediation, finding new ways of representing and applying these to augment our repertoire for these otherwise not-representationally-mediated activity by representational control. And I suggested that, in so doing, attempts to place limits on our repertoire escape our grasp.

But this argument depends on the idea that we can repeat “higher-orderfying” moves to expand our repertoire *ad infinitum*, so it does not clearly go through unless we have a potentially unlimited span of time to do this iterating. Perhaps individual human beings have this same self-developing capacity. But isn’t there a limit to how much self-developing anyone can do in one lifetime?

The answer is that, though our time on earth is limited, what we can do with it depends on the way in which we are ensconced in a broader social setting. We are organized around participating in collective projects, availing ourselves of our collective purposive organization. And we collectively are organized around drawing on the efforts of each of us. The collective version of Universality entails the plan that we individually do ρ is in our collective repertoire: getting you to do ρ is in our collective toolkit – or at least, we can’t rule it out.

The question of whether any such plan is in our corresponding individual repertoire, then, depends on whether the prospective performance of ρ that we collectively have prospects of getting someone to engage in normally can involve *them* behaving normally in doing so: whether it is a matter of engaging their capacities to participate in group activities, or of simply pushing them around. But the Collective Universality Thesis gets us the answer we want. Because the Collective Universality Thesis also covers the plan that the collective gets individual in question does ρ *by means of engaging the individual’s collaborative capacities*.

§2 - The Resulting Picture

2.1 – Truth Normativity

I have given an argument for the Universality Thesis as a characterization of our individual repertoires. At the close of the previous chapter, I gave a preliminary indication of how the Universality Thesis can give us the three conditions for the truth requirement on belief. These are, first, that there be a point to believing P: that among any subordinating set of possible demands we could discharge by believing P, there is at least some one that we have prospects of doing in just this way. Second is that there should be prospects of believing P normally in the first place, that it be possible to get the true belief, and not merely by accident, but in keeping with the normal course of functioning by which you form beliefs. And third, it should be possible to do this – and in particular to do so by means of recognizing an instruction to do so, in a way that avoids conflict with other plans around which you are organized. Now that it is on the table *why* I think the Universality Thesis is true, I would like to say a little bit more about how it comes out that on the resulting picture, the norm to believe P just in case it is true obtains for us.

For each of the three conditions on being called-for, there are cases in which believing the truth would seem, *prima facie* not to be capable of meeting that condition. Some beliefs do seem, on their face, to be pointless to have because they are trivial or otherwise un-actionable – for instance, a true belief about the number of times the letter ‘e’ occurs in this dissertation. Others seem like getting them would conflict with our other aims – because they would take time or effort to obtain or maintain that would be better spent otherwise, or because, given other reasonably held, but false beliefs, they would seem to recommend actions that are in fact counterproductive. And some beliefs seem impossible to get in keeping with normal functioning, because, though they are true, the evidence available misleads in the other direction.

The basic response to all of these concerns is to note that they rely on assumptions about limitations in our repertoire. A piece of information might appear trivial because we cannot think of a contingency plan worth pursuing that depends on it. But, I have argued, there is no telling what endeavors we might have occasion to pursue, and so what information they may require. If all you had to look at what sorts of plan-involvements we might get into left to our own devices, based on the precedents around which a human brain evolved, we could place limits on what we might get up to within our lifetimes. But we are organized around and contribute to projects that span well beyond our lifetimes, ones whose direction we cannot predict. With this in mind, we cannot rule out being surprised about the usefulness to posterity of some piece of information we might now gather. In this connection, it is important to remember, that the open-ended ability to ascend to a new level of symbolic control means that our projects involve not just forming representations that effect action, but also ones that help us to form other representations: beliefs may have their cash value in action at many levels of remove, being more directly related to usefulness in forming other beliefs.

Likewise, concerns about the ability to obtain and maintain a belief without compromising on other goals, by taking up too much time and effort, or by distracting, confusing, or distressing ourselves, revolve around assumptions about limitations on the *ways* in which we are capable of holding or entertaining the beliefs: of being prepared to adapt contingency plans we are trying to further by instruction. But our methods of obtaining and maintaining beliefs are subject to change based on the progress of the purposive activity of the society in which we are ensconced. For instance, there are beliefs about numbers that are not tractable to hold if all you have available to you is a system of notation like roman numerals: trillions can only represented with a sequence of one million 'M's. In "The Sand Reckoner," Archimedes sets out to show that there is a number greater than the number of grains of sand that it would take to fill the universe. It may be hard to see today why this would be doubted in the first place today, but in a situation in which the tools have

not been invented to represent very large numbers, the idea that such a quantity is finite might be hard to concretely grasp. Archimedes resolves this issue by devising a method of expressing very large numbers, reaching quantities as high as $10^{(8 \cdot 10^{16})} = 10^{80,000,000,000,000,000}$, far beyond the quantity needed to express an upper bound on the number of grains of sand that can fit in the universe (Archimedes/Newman, 1956).

Beyond innovations of notation, there are innovations in the underlying means by which we can store and access information: access to information stored on a computer prepares us to act as an informant relaying it (cf. Clark and Chalmers, 1998).

On the question of misleading evidence, as well as on the question of true beliefs that could be misleading for action: again, the intuition that these are not attainable in a way that would be normal or promote downstream normal functioning depends on the assumption that the way of holding the belief most clearly at your disposal would not be normal, because it would be in the face of evidence, or would not be beneficial, because it would combine with your mistakes or ignorance to get you into trouble. But it need not follow that any ways of holding the belief are off the table, because we cannot rule out of your repertoire your getting better evidence or alleviating the mistakes that would make your true belief damaging. It is *this* prospect of having a true belief, not that of getting it in the easy, bad way, that is called-for in such a case.

We are left with a picture on which there is always, as far as we can tell, a possibility for believing the truth in a way that is useful, and deconflicted, and a non-accidental, normal success. Nonetheless, don't the long odds of this kind of success mean that sometimes it is better to pursue not believing the truth anyway? Well, this all depends on the cost of *trying* for the relevant benign option. I believe this trying is free, for us. It is a matter of being open to the better reason, to being convinced by what one takes to be convincing evidence.

The use of the kind of explicating move that it would be useful for Helen to have available, in order to say to the Hare what it is that she is trying to get him to do: “*believe* what I have shown you is *true!*” is a kind of reinforcing effort, a calling-for what the Hare is anyway engaged in doing in thinking at all. It is this sort of explicating reinforcement that counts as the conflict sensitive imperative signal that figures in the definition of called-for-ness. The Hare’s taking in such a signal does him no harm because it is a reminder to him to do what he is inevitably doing anyway, in taking the world to be a certain way.

This idea about inevitability of such an effort – that we cannot but try to believe the truth - is sometimes thought to be objectionable as part of an argument for any kind of truth-directedness of belief, let alone the strong kind of truth normativity I am urging here. The worry is that merely not being able to help doing something doesn’t support the claim that it is the right thing, that some other alternative would not be better if we could pursue it. The trouble with this objection is that to it is one thing to be able to contemplate but unable to achieve some alternative way of proceeding; but it is quite another thing – and, I think, this latter is our situation. We can imagine, from a third person perspective, a creature that did not try to believe what was true, as far as they could, in the way that we do. But if we are contemplating what we will try to do, we cannot imagine from within what it would be to take convincing evidence of the truth of some claim as anything other than compelling reason to be convinced of it.

2.2 – Intentionality, Sociality, and Normativity

It is noteworthy that on the picture of person-level intentionality I am advancing, there is a dual role for sociality in instituting the normativity that I am urging is essential. A central theme of Brandom’s *Making it Explicit*, is that being subject to such to interpersonal enforcement – for him, the key element is the *sanction* – which separates the fact of the deed from the assessment of its

propriety is a key strand of the normativity that he construes as essential to intentionality. This dimension of sociality is, in his words “I-Thou” sociality: it is essentially a matter of specific interpersonal interaction. In a critical essay, Allan Gibbard (2010) asks whether this kind of sociality need be essentially inter-agential, or whether it may be a matter of a creature being able to direct sanctions at herself. Brandom’s reply is concessive (2010 299):

Now any situation that that admits of a distinction of perspective [between that of attributing and that of acknowledging a commitment] is going to count as having a social structure in my “I-thou” sense... Gibbard considers a solitary orangutan, and asks whether *memory* could not suffice to underwrite the distinction of normative attitudes that I insist on. Couldn’t the orangutan *attribute* commitments to its *past* self, and distinguish those from the ones it *now* acknowledges? I do not know whether orangutans can do that, nor whether if as a matter of fact one has to interact with others of one’s kind in order to master the distinction of normative attitudes; that is a broadly empirical requirement or [*sic*; should be “not”?] a conceptual one. But *if* creatures *can* take up the different perspective to time-slices of themselves, *then* the relation among those time-slices is *social* in my sense. For I am *only* claiming that intentionality must be social in the sense that it must admit of the distinction of perspectives between the attitude of attributing a commitment (or other normative status) and the attitude of acknowledging it.

Brandom’s insistence on the sociality of our intentionality, though much discussed, is not as incorrigible as it is sometimes portrayed.

My account provides another dimension of sociality for our intentionality, one on which is also essentially related to normativity, but is not even potentially dispensable in the way that Brandom concedes his is. On my view, the social plays an essential role, not just in its contribution of a specific kind of normatively laden *interaction*, but in the extension of our powers that ensconcement in an indefinitely self-developing social whole effects for us, which allows us to be normatively directed at truth itself, not just at truth-insofar-as-it-is-useful, which repertorial limitations would reduce us to.

2.3 - Into the Unknown

An important issue, one that I have put off addressing until now, arises from the extent to which it relies on possibility, and on a negative epistemic claim about possibility at that. Our repertoire circumscribe what it is possible for us to do, as opposed to merely have happen to us. And the universality claim is the claim that, as long as we cannot rule out the idea that it is possible for something to merely happen to us, we cannot rule out the possibility that it is possible for us to do it on purpose (in the weak sense of “on purpose” in which we grew our teeth on purpose). There is a worry here that this position is quite precarious: it depends on what is possible for all we know, but, at least arguably, a lot less is in fact possible than is possible for all we know. From here, we can’t rule out finding conflict-sensitive ways of holding an arbitrary belief or of making use of it. But who is to say that we couldn’t discover that much less is possible in this regard than what we merely can’t rule out.

Here, my response is simple. Following Sellars (1962), I think a core element of being a person is conceiving of ourselves and one another as persons. And to be sure, there is plenty that we already know it is unlikely that we will ever be able to believe in keeping with our normal functioning or to make use of. But on view I’m advocating, there is all the difference in the world between what is unlikely and what is impossible, because, recall, trying for the relevant unlikelihoods is cost-free. And if we were able to prove that this unlikelihood this would make going on as Helen would like to go on, when equipped with the right vocabulary, in talking to the hare incorrect, and it would make the kind of thinking that this way of going on in speech would make explicit incorrect. In such a case, there are two possibilities.

First, our newfound knowledge about absolute limitations might not come along with insight into another way of going on besides taking truth as compelling grounds for belief. In such a case,

we would simply be confined to the fate of living out a conception of ourselves which we knew to be false – Manifest Truth-Normativity would still be true, but Truth-Normativity would not.

Second, this newfound knowledge might come along with the development of some other way of proceeding which, as yet, we cannot contemplate. In such a case, our enacted conception of ourselves would change radically. And because our enacted conception of ourselves is at the core of what we are, such a change would constitute us becoming something new, something other than (whether “beyond” is appropriate is another question) a person. I think there is no telling what that something would be.

Bibliography

- Archimedes. 1956. *The Sand Reckoner*. In J. Newman (ed.) *The World of Mathematics Vol. 1*. New York, NY: Simon & Schuster.
- Artiga, Marc. 2021. Biological Functions and Natural Selection: A Reappraisal. *European Journal for the Philosophy of Science* 11.
- Bergman, Karl. 2023. Should the Teleosemanticist be Afraid of Semantic Indeterminacy? *Mind and Language* 38/1: 296-314.
- Bigelow, J. and Pargetter, R. 1987. Functions. *The Journal of Philosophy* 84: 181-96.
- Block, Ned.
- 1978. Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science*. 9: 261-325.
 - 2002. Some Concepts of Consciousness. In D. Chalmers (ed.) *Philosophy of Mind: Classic and Contemporary Readings*. New York, NY: Oxford University Press. 206-19.
- Boorse, Christopher. 1976. Wright on Functions, *The Philosophical Review* 85/1: 70-86.
- Brandom, Robert.
- 1994. *Making it Explicit*. Cambridge, MA: Harvard University Press
 - 2000. *Articulating Reasons*. Cambridge, MA: Harvard University Press.
 - 2010. Reply to Allan Gibbard's "Thought, Norms, and Discursive Practice". In B. Weiss and J. Wanderer (eds.) *Reading Brandom: On Making it Explicit*. New York, NY: Routledge. 297-301.
- Carroll, Lewis. 1895. What the Tortoise told Achilles. *Mind* 14/4: 278-80.
- Chomsky, Noam. 1980. *Rules and Representations*. New York, NY: Columbia University Press.
- Clark, Andy, and David Chalmers. 1998. *The Extended Mind*. 58/1: 7-19.
- Cummins, Robert, 1975. Functional Analysis. *The Journal of Philosophy* 72: 741-65.
- Darwin, Francis. 1887. *The Life and Letters of Charles Darwin* (3 vols.). London: John Murray.

- Dawkins, Richard. 1986. *The Blind Watchmaker*, New York: Penguin.
- Davies, Paul Sheldon. 2001. *Norms of Nature: Naturalism and the Nature of Functions*. Cambridge, USA: MIT Press.
- Dennett, Daniel.
- 1989. *The Intentional Stance*. Cambridge, MA: MIT Press.
 - 1995. *Darwin's Dangerous Idea*. New York, NY: Simon & Schuster.
- Fodor, Jerry. 1992. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- Fodor, Jerry, and Massimo Piattelli-Palmarini. 2011. *What Darwin got Wrong*, New York: Polity.
- Garson, Justin.
- 2015. *A Critical Overview of Biological Functions*
 - 2017. A Generalized Selected Effects Theory of Function, *Philosophy of Science*, 84/3: 523-43.
 - 2019. *What Biological Functions are and Why They Matter*. Cambridge University Press.
 - Forthcoming. The Story of Functions. *The Australasian Philosophical Review*.
- Gibbard, Allan. 2010. Thought, Norms, and Discursive Practice. In B. Weiss and J. Wanderer (eds.) *Reading Brandom: On Making it Explicit*. New York, NY: Routledge. 15-32.
- Godfrey-Smith, Peter. 1994. A Modern History Theory of Functions. *Noûs* 28/3: 344-62.
- Gould, S.J. and Lewontin, R. 1979. Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society B*405: 581-98.
- Gould, S.J. and Vrba, E. 1981. Exaptation: A Missing Term in the Science of Form. *Paleobiology* 8: 4-15.
- Gould, Stephen Jay. The Exaptive Excellence of Spandrels as a Term and Prototype. *Proceedings of the National Academy of Sciences* 94/20: 10750-5.

- Griffiths, Paul. 1993. Functional Analysis and Proper Functions. *The British Journal for the Philosophy of Science* 44/3: 409-22.
- Haig, David. 2020. *From Darwin to Derrida*. Cambridge, MA: MIT Press.
- König, Claus and Weick, F. 2008. *Owls of the World* (2nd ed.). London, UK: Christopher Helm.
- Lennox, J.G. 1993. Darwin was a Teleologist. *Biology and Philosophy*, 8: 409-21.
- Longy, Françoise. 2009. How biological, cultural, and intended functions combine. In U. Krohs & P. K. Kroes (Eds.), *Functions in biological and artificial worlds: Comparative philosophical perspectives* (51–68). MIT Press.
- Machery, Edouard. 2012. Why I Stopped Worrying About the Definition of life... and why you Should as Well. *Synthese* 185/1: 145-64.
- Moreno, A. and Mossio, M. 2015. *Biological Autonomy: A philosophical and Theoretical Enquiry*. The Netherlands: Dordrecht.
- Mossio, M. and Bich, L. 2017. What Makes Biological Organisation Teleological? *Synthese* 194/4: 1089-1114.
- Millikan, Ruth Garrett.
- 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
 - 1995. *White Queen Psychology*. Cambridge, MA: MIT Press.
 - 2002. Biofunctions: Two Paradigms. In A. Ariew, R. Cummins, and R. Perlman (eds.) *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press. 113-43.
 - 2005. *Language: A Biological Model*.
- Nanay, Bence. 2010. A Modal Theory of Function. *The Journal of Philosophy* 107/8: 412-31.
- Neander, Karen.
- 1991. Functions as Selected Effects: The Conceptual Analyst's Defense, *Philosophy of Science* 58/2: 168-84.
 - 2017. *A Mark of the Mental: In Defense of Informational Teleosemantics*. Cambridge, MA: MIT Press.

- Paley, William. 1963 (1802). *Natural Theology*. Indianapolis and New York: Bobbs-Merrill.
- Papineau, David. 1993. *Philosophical Naturalism*. Cambridge, MA: Blackwell.
- Rubner, Andrew. 2023. Mechanistic Explanations and Biological Functions. *The British Journal for the Philosophy of Science*.
- Sellars, Wilfrid.
- 1954. Some Reflections on Language Games. *Philosophy of Science* 21/3: 204-28.
 - 1960. Philosophy and the Scientific Image of Man. In R. Colodny (ed.) *Frontiers of Science and Philosophy*. Pittsburgh, PA: University of Pittsburgh Press. 35-78.
 - 1968. *Science and Metaphysics: Variations on Kantian Themes*. Atascadero, CA: Ridgeview.
 - 1981. Mental Events. *Philosophical Studies* 39/4: 325-45.
- Sosa, Ernest.
- 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1*. New York, NY: Oxford University Press.
 - 2011. *Knowing Full Well*. Princeton, NJ: Princeton University Press.
- Sperber, Daniel. 2007. Seedless grapes: Nature and culture. In E. Margolis & S. Laurence (Eds.), *Creations of the Mind*. 124–137. Oxford University Press.
- Watson, Adam. 1957. The behaviour, breeding and food-ecology of the Snowy Owl *Nyctea scandiaca*, *Ibis*. 99/3: 419–462.
- Wittgenstein, Ludwig. 2009 (1953). *Philosophical Investigations*. Tr. G.E.M. Anscombe, P.M.S Hacker, and J. Schulte. Malden, MA: Wiley-Blackwell.
- Wright, Larry. 1973. Functions, *The Philosophical Review* 82/2: 139-68.