Group Agency and Biological Individuality

David Kelley

A thesis submitted in fulfilment of the requirements for the degree of Doctor in Philosophy, the University of Auckland, 2022.

Abstract

This thesis explores the possibility that the topic of group agency (in moral and political philosophy) and the topic of biological individuality (in the philosophy of biology) can be mutually informative. I argue that ideas and perspectives germane to one topic might be theoretically useful when taken up in the context of the other. There is a natural perspective we take towards humans and human groups, that if applied to our theorizing about biological individuals, allows us to give fuller explanations and descriptions of certain cases in which philosophers of biology have taken an interest (such as those in which individuals at different hierarchical levels exhibit individuality to a degree). Conversely, there is a pluralistic approach commonly taken toward biological individuality that, if applied to cases of group agency, suggests a useful, more nuanced taxonomy that distinguishes different kinds of agents or degrees of agency. Beyond what is argued in detail, I have made efforts to raise a number of related questions and problems, outlining them for future work in this interesting novel intersection of subfields.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Tim Mulgan and Emily Parke. Anything that shines in this work will be a credit to their efforts.

Table of Contents

Introduction

This thesis explores how two areas of philosophy might be mutually informative. The two topics of concern are group agency within moral and political philosophy and biological individuality within the philosophy of biology. What does it mean to be "mutually informative?" Minimally, this means that one topic has the potential to inspire novel inroads for the other topic due to certain analogical concepts, however deep or superficial. I believe this is true. But it is also true about a great many topics.

A more specific hope lingering behind the phrase "mutually informative" may be theorized in terms of supplementing perceived deficits. For example, we might hope that generally, philosophical discussion of group agency could benefit from some of the empirically driven insights of philosophy of biology. In return, perhaps philosophical discussion of biological individuality could benefit from the existing theoretical work on group agency.

What motivates conversations about group agency and biological individuality is wanting to get our descriptions of the world right and preserve what matters when we act on those categories. "What matters" may refer to the things categorized (e.g. conscious beings) or refer to what matters to us *in the categorizing* (e.g. kinds of agents or individuals that feature in our descriptions and explanations). What does it mean to "get our descriptions right?" We miss something when we don't see a honey-bee colony working as, and selected as, a unit. We miss something when we don't see powerful corporations as unified actors making decisions and carrying out plans that affect millions. Given that group agents are real only as social constructions, the group agency literature is bereft of any empirical, real-world givens. Therefore, we might want to look to the natural world to shore up certainty about our chosen concepts and models. That is, the plausibility of group agents has to be argued from the ground up. This is not so in the case of biological individuals, which we take to exist in nature independent of our theorizing about them. However, the biological individuality literature arguably lacks simple, consistently applied theoretical tools to delineate biological entities in a way that preserves the details we care about in some cases (e.g., in the interactions between different kinds of individuals or individuals at different "levels").

To be sure, these kinds of motivations and expectations can be overstated. I proceed largely in the spirit of philosophical exploration. Having said that, this thesis culminates by offering specific ways forward for subtopics of potential interest. Thus, I will conclude that indeed,

these two subfields can be mutually informative.

A proper introduction to each of the topics concerned - group agency and biological individuality - will be provided in chapters 1 and 3 respectively. But very briefly, for the sake of the rest of the introduction that follows, here is a snapshot of these two subjects.

We often talk as if groups are agents, so it is important to distinguish talk from reality. Unlike mere groups, group agents are agents in their own right, not reducible to the agency of their members. What largely motivates philosophical accounts of group agency is a need to hold group actors responsible for their actions, especially when no individual is at fault.

Biological individuality is not straight forward. While paradigm individuals are easy to spot (e.g. kangaroos), much of the living world is not so intuitively distinguished. What motivates philosophical accounts of biological individuality is our natural desire or practical need to be able to count individuals, track individuals over time, and to distinguish organisms from their detritus and offspring.

Two Literatures - Two Approaches

There exists a substantial philosophical literature on the topics of group agency and biological individuality. One option for exploring either subject would be to survey a range of accounts and group them into various types. For example, one might group accounts by how liberally the authors apply the status of "agent" to groups – how much daylight falls between mere groups and group agents; or in the case of biological individuality, group accounts according to the kinds of features under consideration as criteria for individuality. Another option is to start with a single clearly argued account, then subject it to perturbation and critique to see how accommodating the account can be towards a range of different *desiderata* expressed in the literature. In the case of biological individuality, I have opted for the former approach. For the topic of group agency, I have chosen the latter route.

In contrast to my approach to group agency which "zooms in" on a single account, I take a "zoomed out" approach to the literature on biological individuality. I survey the literature more broadly, briefly featuring a few important accounts. This approach is most suitable given my focus on the current trend toward pluralism about individuality (which is a feature of the literature generally, not a single account). I examine a range of case studies which put

pressure on the pluralist's ability to account for and fully explain these cases with (even multiple) individuality concepts alone.

My approach to group agency requires a bit more explanation. Of all the existing accounts of group agency, none prioritizes group-level rationality to the extent that List and Pettit do, as featured in their book *Group Agency*. Their emphasis on a decision procedure aims to demarcate rationally unified group agents from less unified groups. Groups qualify as or become group agents when rational unity is guaranteed through group structures and procedures that ensure rational failures cannot occur. List and Pettit's focus on rationality is arguably the most suitable starting point for exploring analogies with the as-if-agency of biological "agents" based on the agential heuristic I adopt. But also, and importantly, the rigidness of List and Pettit's agent-constitutive requirement for group-level rationality, combined with their under-explored notions of "inexplicit decision procedures" and non-intentional group formation, makes them a good target for less clear cases. I will apply pressure to List and Pettit's account making use of such cases (5.1), challenging the consistency of the demarcations they offer (2.3), and evaluating aspects of their account in view of how others think about agency (2.2).

Thus, while I have drawn on considerations from other accounts of group agency, prioritizing the work of List and Pettit is a project-constraining commitment. This thesis does not attempt a systematic engagement with the wider group agency literature. Rather it prioritizes List and Pettit's account by presenting and critiquing parts of it in more detail than a work of this type could normally accommodate. Similar projects could be done prioritizing other prominent works.[1]

Outline of the Thesis

In what remains of this short introduction, I will outline the plan for the thesis, briefly summarizing the aims and contents of each chapter.

In Chapter 1, I lay out List and Pettit's account of group agency. Because I aim to mostly present the parts of their account in a chronology similar to that within their book, this

[1] For example, Raimo Tuomela's *Social ontology: Collective intentionality and group agents* (Oxford University Press, 2013).

treatment will start with the basics: what are group agents, as opposed to groups? What motivates accounts of group agency? What is an agent anyway? The chapter gives special attention to the rational unity of groups and the examples which show how this may be achieved. Latter portions of the chapter that features List and Pettit's account of group personhood which will lay the groundwork for criticisms developed in chapter 5.

Chapter 2 provides further context for List and Pettit's account. I start by featuring precursors to List and Pettit's account - those that have laid the foundation for and motivated List and Pettit's work, namely Peter French and Seumas Miller. Next, I consider contrasting accounts of agency more generally. The conceptions of agency featured in the work of Carol Rovane and Christine Korsgaard both put pressure on List and Pettit's conception of agency and lay some important groundwork for connections with the biological individuality literature. Building on the previous sections, I will offer a critique of relevant aspects of List and Pettit's account of agency. Lastly, I will feature work by Stephanie Collins, who offers a taxonomy of different types of collectives. Each of these sections prefigure unsatisfactory aspects of List and Pettit's account for which I will offer possible solutions, as inspired by the topic of biological individuality, in chapter 5.

Chapter 3 provides a literature review on the topic of biological individuality. After laying out the main problems, I feature two leading accounts – those of Ellen Clarke and Thomas Pradeu. Following this, I feature the work of Maureen O'Malley and consider some specific cases of candidate individuals that put pressure on our intuitions on the subject and motivate the case for pluralism. I then feature three pluralistic approaches to biological individuality, those of Peter Godfrey-Smith, Jack Wilson, and James DiFrisco. This chapter lays the groundwork for an argument that agential descriptions and language are useful independent of whether and when various biological entities count as individuals of one kind or another, and that keeping the concept of agency front and center is useful when theorizing about biological individuals. In Chapters 4 and 5, I will show that even pluralism about biological individuality is sometimes insufficient or unsatisfactory to provide the fullest, best explanation of various individuality phenomena (e.g. puzzling marginal cases). For better explanations, I will go on to argue, we need to pull from a perspective that comes naturally when engaging with the group agency literature.

I spend the first half of Chapter 4 defending the use of teleological language in biological explanations. I then introduce the work of Samir Okasha who extends the utility of similar

heuristic theorizing to include talk of agents and goals. I argue that agential unity, as achieved by compatible intentional states or biological traits, can serve as common currency between the topics of group agency and biological individuality.

In Chapter 5, I offer my original contribution to the subfields of group agency and biological individuality by suggesting ways in which each can usefully inform the other. In the first part of the chapter, I suggest that regardless of whether a pluralistic approach towards biological individuality is appropriate or satisfactory, it nevertheless offers a powerful lesson for group agency, particularly List and Pettit's account. Specifically, I suggest that a novel or revised taxonomy would help preserve the distinctions that matter for List and Pettit, those not carefully enough distinguished in their own account (namely, agents who do and do not possess morally commanding interest). I further argue that additional nuanced taxonomies are possible and potentially useful, featuring types of different agents suited to our different purposes, not unlike the different senses of "biological individual" surveyed in chapter 3. I then offer a multi-dimensional model for thinking about different types of collectives – a model adapted from a pluralistic account of biological individuality in which individuality admits of degrees.

In the second part of the chapter, I highlight a perspective we naturally take when considering group agents, "the multi-level agential stance," and how it can refine and strengthen our application of agential heuristics in our explanations of various individuality phenomena. Specifically, I suggest we can shift levels of analysis without shifting stances, just as we do in the case of group agents in which we naturally think of the parts of an agent as themselves agents (i.e. individuals within the group). By exploring specific cases (e.g. cancer), we can see how better explanations follow from agential descriptions regardless of whether the agents we describe qualify as individuals or not. The reader is then offered a tour of multi-level phenomena featuring cases on which to practice shifting stances (e.g. altruism, parasites). I then introduce the term "methodological agentialism" to denote the approach of deploying agential heuristics while in the multi-level intentional stance.

In the short third part of the chapter, I offer an aspirational glimpse of the kind of ongoing synergy these two subdisciplines might create by way of further engagement in the style of my thesis. Specifically, following Okasha, I show how pareto violations in social choice theory can inform the literature on evolutionary transitions, which in turn can feed back into the group agency literature, shoring up List and Pettit's theoretical criterion with empirical

biological facts about groups and individuals. The chapter concludes on the hopeful note that this may be just one of many instances in which we recursively strengthen the bridge between the two subfields.

Chapter 1
List and Pettit's Account of Group Agency

In this first chapter, I will present relevant portions of work by Christian List and Philip Pettit (hereafter, LP), mostly from their co-authored book, *Group Agency*. The scope of this thesis is mostly limited to LP's account of group agency, giving short attention to relevant ideas from other accounts in chapter 2. Section 1 introduces the subject of group agency by calling attention to the way we talk about different groups. Section 2 lays out LP's criteria for minimal agency. Section 3 and 4 cover group intentions and group formation. Section 5 presents LP's trademark illustration of group level rationality. Sections 6 and 7 discuss desirable qualities for group agents, group responsibility, and group personhood.

1.1 Speaking of Groups as Agents

In their book, *Group Agency*, LP argue that it is possible for groups of individual agents to comprise a group agent. Their position is that a group of individuals with the right sort of structure and decision procedure can qualify as an autonomous, rationally unified agent in its own right – possessing an agency distinct from ("over and above") the agency of its members.[1] The agency of such groups should be understood as real and non-redundant. Some groups can be "centers of attitude and action," i.e., agents.[2]

LP begin their investigation of group agency by analyzing the way we talk about some groups. We often refer to a group of individuals in a way that suggests they are not just a collection but also a unified whole. We can consider how to variously understand statements that do so:

1. The group is really a unified whole.
2. Treating the group as a unified whole is useful for explanations and predictions, but the group is not really a unified whole.
3. Referring to the group as a unified whole is an artifact of linguistic convenience (i.e. our use of language *groups*, but does not *refer to* a unified group).

[1] Christian List and Philip Pettit, *Group Agency*: *The Possibility, Design, and Status of Corporate Agents* (Oxford: Oxford University Press, 2011), 78.
[2] List and Pettit, *Group Agency*, 194.

LP aim to lay out criteria for legitimate cases of #1, in which the unification of the group is of the kind that makes the group a center of attitude and action, i.e., an agent.

Let us explore a common motivation for considering agency at the group level by noting a few specific examples of how we speak about groups such as clubs, firms, and states. Consider the following statements as they might be encountered on a daily basis:

The firm intends to double their efforts.

Japan is worried about its aging population.

The judicial panel doubted the defendant's testimony.

The Philip Morris Company doesn't care about anyone.

Victoria University prefers to hire professors from overseas.

Apple believes they will face more competition in the coming years.

China vows to retaliate over the U.S. president's threat of new trade tariffs.

Lodge 234 feels that use of Masonic property should be reserved for members only.

The Catholic Church has promised reform in response to cases of sexual misconduct.

We can make sense of these statements with little effort. In each case, we immediately identify the group referenced and understand the way in which the intentional idioms employed are useful and informative. Given that this way of speaking and understanding is so common and comes so naturally, we might wonder if these statements refer in a literal way, or simply reveal a psychological predilection for anthropomorphizing groups. For LP, the answer turns on the type of group – and in the cases above, all the groups referred to would be candidate group agents.

We sometimes encounter similar kinds of sentences that pick out a mere grouping of people based on some characteristic: *light-skinned people prefer the shade*.[3] The collection of light-skinned people is obviously not a group agent (and it is doubtful any utterer of this sentence takes them to be). Similarly, we should expect that a headline reading, *Train Riders Prefer Higher Fares to Longer Wait Times* is reporting the summed results of a survey taken by train riders, not that all train riders make up a unified agent that holds this preference. Therefore, sometimes this manner of speaking is just shorthand, a linguistic act of *grouping* for the sake of semantic convenience (an instance of #3 above). By contrast, with the list of statements above, it is clear that in making such statements we are not merely *grouping*, but referring to groups as actors, desirers, believers, or decision makers (e.g. Apple, having assessed and judged that the market will produce strong competitors, believes they will face more competition in the coming years).

Sometimes we do posit agency as a heuristic methodology. It is this understanding of some sentences that bridges the realm of metaphor and realism. If these kinds of explanations and predictions consistently prove to be informative and useful ways of reasoning (i.e. heuristics), then we might count such usefulness as evidence that the observed system really is an agent. Indeed, in the statements listed above, it seems likely we would make use of such a heuristic, for example, in making sense of Chinese-American foreign relations, say, if America proceeded with imposing such tariffs and then China were to retaliate by imposing its own tariffs, and so on. Daniel Dennett refers to what I have described here as taking the "intentional stance" – in this case, treating the group as a goal-directed intentional agent and understanding the group's decisions, actions, and statements in terms of beliefs and desires.[4] However, LP will not allow agency to be ascribed based on this utility alone – certain conditions within the agent must be met. The latter point will be a key difference between competing accounts of group agency.[5] For LP, the consistent utility we may get from treating a group as an agent is not sufficient for ascribing real agency, but rather is a motivation for taking seriously a candidate group agent. (We will later return to Dennett's view which has an important role to play in chapter 5 of this thesis).

[3] For LP, such collections of individuals should not even be thought of as "groups," much less group agents; they are "mere collections" (31).
[4] Daniel Dennett, *The Intentional Stance* (Cambridge: MIT Press, 1987). See also, List and Pettit, *Group Agency*, 23.
[5] e.g. Functionalist versus interpretivist accounts of agency (LP's is a hybrid).

In some cases, our ways of speaking will not be metaphorical, and yet will still not pick out group agents; sometimes group-agency talk is meant literally but is inaccurate or misguided in some way. Someone might talk of the market, conceiving of it as an agent in its own right – one that we can track, think of as undergoing changes over time, and make broad predictions about. Upon closer examination, it is clear that a market could not, for example, enter into a contract (like a group agent could).[6]

Other times, we may fail to pick out group agents, not because the thing we indicate is not the kind of thing that could be an agent (e.g. the market), but because of vagueness due to ignorance, often couched in metaphor, as with metonymy. Consider citing the "the heads of state" or "capitol hill" as the cause of some action or utterance. This circumscribes a collection of agents that may contain the causally responsible agents. It may be preferable to talk in this vague way when we do not know the precise locus of attitude and action, and can only approximate it. In such a case, we may have picked out a group agent, but we have over-attributed membership in a way that is imprecise and therefore of little use, and as a result we are not able to carefully track the attitudes and actions of the same group agent.[7]

However, other times, talk of groups as agents will neither be mere metaphorical nor misguided, nor merely a heuristic – sometimes such talk will in fact pick out group agents. LP argue that in some cases where talk of certain groups is not easily translated into individual terms, we can think of such groups as "autonomous realities."[8] That is, we are *justified* in our reasoning about them as agents – *justified* in our taking the intentional stance towards such groups. We are justified because those groups *really are agents*. LP present certain criteria groups must meet to count as agents, and thus for such attributions of agency to correctly refer.

To summarize LP's position on talking about groups as agents, we often refer to groups as if they were intentional agents. Sometimes speaking this way is intended simply as metaphorical shorthand. Other times, speaking this way might be intended literally, but mistaken. While either of these two uses may be a succinct and useful use of language, they are not correct should we take them to be picking out groups that are real unified agents.

---

[6] Philip Pettit, "Group Agents Are Not Expressive, Pragmatic or Theoretical Fictions," *Erkenntnis* 79, no. S9 (2014): 1644.

[7] It should be noted that the use of metonymy and other figures of speech do not necessarily imply that an instance of group-agency talk is metaphorical. However, it may result in ambiguity in cases of nested group agents.

[8] List and Pettit, *Group Agency*, 6.

However, in some cases, referring to groups as intentional agents, will not only be literal, it will be correct, as well as useful (perhaps even indispensable or unavoidable). LP will conclude that recognizing these groups as real agents is both in society's best interest and consistent with our best practices.

1.2 Minimal Agency

Our conviction that we are dealing with an intentional system is derived from the explanatory power provided by treating it as such.[9] While the explanatory power of our practices is enough to motivate entertaining the notion of group agency seriously, LP aim to go further, undertaking the task of laying out criteria that *must* be met for a group to really count as an agent. In order to make their case for the possibility of group agency, LP first lay out the minimal conditions that must be satisfied by all agents. For LP, all agents must have at least these three features:

1. Representational states
2. Motivational states
3. The capacity to process and act on the basis of those states.[10]

The first feature corresponds to what, in the case of an individual human agent, we would call "belief." The second feature corresponds to what we would call "desire." Features 1 and 2 are intentional states that play the role of depicting the world and motivating action, respectively. Intentional states are functional, meaning that they are defined by the role that they play, not by any particular realization.[11] We normally consider these states in a context in which they are realized in the physical form of a human brain. However, no particular physical instantiation is required of a functional kind. For example, consider a key. A key is that which unlocks something. A key is no more or less a key whether it be a piece of metal, a sequence of numerals, or an RFID chip. Functional kinds should not be confused with broad classes of things. Saying that a knife and a gun are both weapons is not the same as saying both computer encryption and padlocks are both locks (both have a locking *function*).

[9] Deborah Tollefsen, *Groups as Agents* (Cambridge: Polity Press, 2015), 103. See also Bruno Mölder, *Mind Ascribed an Elaboration and Defence of Interpretivism* (Philadelphia: John Benjamins, 2010).
[10] List and Pettit, *Group Agency*, 20.
[11] List and Pettit, *Group Agency*, 21.

LP theorize that these intentional states consist of an intentional attitude held toward a proposition.

The agent believes the house is clean.

The proposition here is "the house is clean." The attitude held toward the proposition is belief.

The agent prefers the house be clean.

The proposition here is "The house is clean." The attitude held toward the proposition is desire.

The third feature is what allows an agent to act. An agent is one that acts on its motivations, guided by its representations. If the agent holds a representational attitude toward the proposition that "the house is messy" and a motivational attitude towards "the house is clean," it will intervene in the environment to make the latter true (i.e. it will clean the house).

According to LP, if a candidate agent possesses these three features, it counts as a minimal agent.[12] The notion of basic or minimal agency motivates an exploration of what sorts of entities in addition to group agents may count as agents (e.g. microbes, robots) given the above three-fold criteria. I will not debate which entities do or do not have representational states; the debate on representationalism falls outside the scope of this project. I am interested in the unity of agency as achieved by the consistent integration of such states (given an agent possesses them). In chapter 4, I will relate the unity of intentional states (representational and motivational) in an agent to the unity of traits in a biological individual.

On LP's account, minimal agents need not have any ability to reason nor make memories or plans.[13] However, they must fulfill standards of rationality at some "minimal level" if they are to count as agents at all.[14] Agents *accurately* represent the world, act *in accordance with*

[12] A "minimal agent," in other words, is an agent by virtue of meeting these "minimal requirements" (5).
[13] List and Pettit, *Group Agency*, 21.
[14] List and Pettit, *Group Agency*, 24.

those beliefs, and maintain coherence of beliefs and desires, avoiding inconsistency and means-end failures.[15] These standards must be met if one is to count as "fully rational."[16]

Violation of these standards can be easily illustrated by considering what it is like for a person to act and speak without regard to such standards:

> Pete is hungry and **wants a sandwich**. Pete decides to act on that desire with the intention of fulfilling it. He goes to the deli, buys a sandwich, and then returns home. He proceeds to unwrap the sandwich and **place it in the trash can** (a violation of what LP call "attitude-to-action").
>
> Seeing Pete gazing drearily out the window at the **sun-drenched** garden, you ask him why he looks so glum. Pete responds, "You know me – I always **feel down when it rains**." (a violation of "attitude to fact").
>
> Later at the bar, Pete confides in you that he is heartbroken because the international charity he runs has failed to meet his earnest expectation to achieve an **above-average** quality of life for **everyone** in the world (a violation of "attitude-to-attitude").

Merely *possessing* rationality does not mean that an agent strives to be rational, nor even ensures that it is aware of its own rationality. An agent may have "no desires or beliefs about its rationality as it updates its beliefs in light of evidence, as they connect with other beliefs and desires, and as they lead to actions."[17] These further capacities go beyond rationality to what LP dub "reasoning." Reasoning, on this account, is an "intentional exercise… taking active steps to form beliefs in metalanguage propositions, so as to check lower-level processing."[18] Therefore, while a mere rational agent may blindly follow its "programming" to act in accordance with its intentional states, reasoning agents can hold intentional attitudes towards the relation between propositions. Instead of, for example, merely holding a belief about an object, a reasoning agent can hold a belief about a proposition.[19]

[15] List and Pettit, *Group Agency*, 24.
[16] List and Pettit, *Group Agency*, 24, 29.
[17] List and Pettit, *Group Agency*, 29.
[18] List and Pettit, *Group Agency*, 30-31.
[19] List and Pettit, *Group Agency*, 30.

Returning to our examples, Pete has clearly acted and believed irrationally. However, we should not be so quick to think he is an irrational individual, that is, that he attempts but *always* fails to meet the standards of rationality. Rational agents *sometimes* behave irrationally. Nor should we dismiss Pete as simply *arational* – that he does not even *attempt* to conform to, and is possibly unaware of, the norms of rationality. In most everyday cases, we would say Pete is likely a rational agent that acted irrationally (as we may observe he is rational in many other ways – he rarely contradicts himself, has reasons from his preferences and beliefs, makes accurate assessments in the work place, and apart from these bizarre instances with the sandwich, rain, and charity, usually gets on with day to day tasks with ease). Reason plays an important role in detecting and correcting irrationality. For example, upon reflection, Pete may come to realize that it does not even make sense that the entire population could experience a quality of life above the mean. Having made clear the incompatibility of these two ideas to Pete, he may then realize there is nothing about which to be heartbroken. Reasoning agents can troubleshoot, revise, and learn. They may fail to always be rational, but any demonstration of sustained efforts to maintain rationality is evidence of one's rational agency, and further, the use of reason (as defined by LP). While the capacity for reason is not a requirement of minimal agency nor even rational agency (though see below), it plays an important role in fulfilling a satisfactory structure for group agents, according to LP.

In addressing the subject of rationality, I should call attention to a curious aspect of LP's account. After laying out minimal conditions for agency, LP move on to introduce constraints and requirements above that of the original definition (the aforementioned capacities of an agent to meet standards of rationality: attitude-to-attitude, attitude-to-fact, and attitude-to-action). Additional capacities are required for aiming to be rational and checking rational processes via meta-language propositions (aka reasoning). Unfortunately, LP make no attempt to carefully articulate successive definitions for the different types of agents or different levels of agency that correspond to the fulfillment of these latter conditions. (Or rather "desiderata," as the minimal conditions for agency have already been satisfied - see sections 1.6, 1.7, and 2.2.3, and 2.3).[20] Concerning the subject at hand, we seem to have a

[20] Although, as I will suggest in chapter 5, more fully fleshing out these sketched categories or levels of agency is a worthwhile task.

requirement for rationality added to the conditions for minimal agency, leaving us to wonder whether an agent is actually a minimal agent or alternatively, if rationality is not truly required for minimal agency. Perhaps this lack of clarity is an artifact of Dennett's influence; on Dennett's view, when we take up the intentional stance, we *assume* that the entity is rational as a starting point.[21]

Maybe we should read "the capacity to process"[22] intentional states as including the ability to avoid inconsistencies and follow the rules of logic (i.e. rationality). I think this is the best way to read LP. Yet, as we shall see, it is LP's notion of *reasoning*, and not mere rationality, that not only realizes these abilities on the group level (as we will see), but also implicates the normative aspect of agency that other philosophers (e.g. Korsgaard, see section 2.2.1) define in terms of an *aim* or *commitment* to being rational (i.e. commitment, not just achievement). Considering LP's emphasis on group level rationality in their treatment of group agents, I will, going forward, take it that LP hold rationality as a requirement for agency.

1.3 Group Intentions

We have seen that, for LP, agency requires a bare minimum: intentional states and the capacity to act from those states. We have also seen that agents must be rational. But how would this apply to groups? We have noted in section 1.1 that we often encounter statements that pick out groups as if they were agents. However, we have yet to answer the relevant question: How can we know *which statements* are figurative shorthand, as opposed to those that actually pick out group agents? How can we tell which groups are set up to meet LP's criteria?

Group agents are usually formed through joint intention, are identifiable over time, and survive changes in membership.[23] They must meet conditions of minimal agency and the standards of rationality. LP demonstrate how groups can be agents by way of how they organize themselves, using decision procedures that partially ensure the aforementioned

[21] List and Pettit, *Group Agency*, 6, 11, 13, 23. i.e., an artifact of the confluence of functionalist and interpretivist accounts of agency, Dennett being in the latter camp.
[22] List and Pettit, *Group Agency*, 20.
[23] List and Pettit, *Group Agency*, 32.

conditions are met. Before we get into this, however, we should consider the formation of groups, a topic which brings up the subject of the *intention* to form a group.

The voluminous literature on collective intentionality, shared intentionality, and joint intentionality is related to, but importantly distinct from LP's account of what it is to be a group agent. I will largely set those debates aside. The important point, for LP, is that group intention is not sufficient for group agency. To be a group agent is not merely (or not even) to intend together.[24]

LP follow Bratman's account of joint intentionality.[25] To say that a group of people jointly intend to do something is to say that each individual forms intentions with, and because of, others who all share the same goal, each intending to play their part with all of these factors being common knowledge among members.[26] However, a group that intends and acts together is not necessarily a group agent. In other words, group agency is not just "joint agency."[27] While random groups of people may spontaneously act and even intend together, group agency requires an intentional formation, a structure that pools the cognitive resources of its members, and upholds consistency in its decisions and proclamations over time.

Before continuing on to the subject of group formation, it will be useful to illustrate how joint intention and collective action do not necessarily add up to group agency:

> There's a horrific car crash on the street beside a motel. The car has tipped over and the driver appears unconscious. John and Dale each rush out of their motel rooms to help pull the driver from the vehicle. Dale yells for Jane to phone an ambulance. Jane immediately calls an ambulance, standing by the car, reporting the details of the rescue in real time. John and Dale lift part of the car to free the driver's hand, which had been trapped. Daisy and Bobby fashion a tourniquet and put pressure on the driver's many wounds as directed by Jane, who repeats the instructions of the emergency operator on the phone. Dale points out that the driver's medical bracelet

[24] LP allow the possibility of non-jointly intentional group agents, e.g., formed by way of convergent evolution, or arranged by some mastermind, e.g. a terrorist managing various independent cells (32-33).
[25] Michael Bratman, *Faces of Intention: Selected Essays on Intention and Agency* (Cambridge: Cambridge University Press, 1999).
List and Pettit, *Group Agency*, 33.
[26] List and Pettit, *Group Agency*, 33.
[27] List and Pettit, *Group Agency*, 32, footnote 18.

had come loose and remains crushed under the heavy wreckage. With her small hands, Sarah is able to retrieve the relevant portion. While still on the phone with the ambulance, Jane hands Sarah a pen, miming instructions to Sarah to write the relevant information from the bracelet plate on the driver's skin so as to be clearly visible to paramedics and doctors in the ER. Meanwhile Derek and Debbie each run to opposite ends of the block anticipating the arrival of the ambulance, ready to wave them from the intersection towards the crash site.

Because of their swift action and efficient coordination, the driver survived. The next day, the local newspaper reports, "Motel Occupants Save Life."

There can be no doubt that saving the driver's life was a joint action. Each person performed their part in rescuing the driver. Many individual actions required cooperation and coordination, such as John and Dale lifting a car they could not each lift on their own. It makes sense to say that the occupants of the motel rescued the driver. That is, they performed a collective action. Note that collective action does not require proximity or physical contiguity; an important role was played by the emergency operator who was not even present. Nor does collective action require previous acquaintance, commitment, or planning. Aside from the couples, the motel's occupants had shared no previous history.

The rescue seems to involve not only joint action, but also joint intention. Each acted towards a common goal of rescuing the driver, with expectations of others doing their part in cooperation. Though the plan was spontaneous, the goal was clear. Each believed the others shared their intentions to cooperate, and that confidence in the intentions of others becomes mutually reinforcing of each's view of the others' beliefs and commitments.

Despite all this, the motel occupants and the emergency operator cannot be considered a group agent on LP's account. There was no decision procedure in place, no attempt to structure themselves so as to prohibit the violation of rational norms, and of course, there was no deliberate, jointly intentional group formation. (As we will see, groups that jointly intend and act merely share a goal, but those that comprise a group agent *share reasons,* whether or not they share a goal).

1.4 Group Formation

Forming a group agent requires that the members of a group "each intend that they together act so as to form and enact a single system of belief and desire."[28] While the collective action illustrated in the example above (in section 1.3) required a substantial amount of coordination and cooperation, what was lacking was this commitment to enact a single system of belief and desire. Though LP do not state it directly (apart from the word "each" quoted in the previous sentence), it seems clear that the group's formation requires every founding member's participation. This is perhaps much the same as we imagine the unanimous agreement of a new political community said to generate correlate pairs of claims and duties (i.e. privileges and obligations) that ensure members remain accountable to the group for their individual commitments to act in accordance with the group agreement.[29] So we can note an important distinction between *group formation* and *group operation*. Once a group agent is formed, not all the members are needed to jointly intend that all actions and procedures of the group be carried out, or even intend that it form certain attitudes. This allows for flexibility and efficiency, permitting a variety of different organizational structures, including division of labor, specialization, sub-units subordinate to other units, etc.[30] Again, this is much like the beginnings of a political community, whose first decision may be a unanimous agreement to institute majority voting. Following this, any number of decisions might be made without unanimity. If a majority then votes for the community to take on a representative form of government in order to facilitate a more efficiently functioning political system, all members are bound to it.[31]

The minimal requirement for group agents is that they are "organized so as to seek the realization of certain motivations in the world and to do so on the basis of certain representations about what the world is like."[32] Group agents will vary in how they are organized depending on their history and purpose. What is important is that a group agent's

[28] List and Pettit, *Group Agency*, 34.
[29] LP tend to avoid this language. Cf. Gilbert's account based on mutual commitments: Margaret Gilbert, *Joint commitment: How we make the social world* (Oxford University Press, 2013).
[30] List and Pettit, *Group Agency*, 35-36.
[31] To briefly point to potential disanalogies: what about children born into established political communities? Some group agents do not face the kind of objections for the analogous case in political philosophy. Non-founding, late-joining members of groups make their own decision to join (or accept a contract, etc.). However, when we consider states and include the populace (not just certain individuals in the government) to be members, this can be problematic. One could reply that this is a non-issue when the age of legal autonomy corresponds with the voting age. One can either remain in the community and have their say according to pre-established terms, or leave the community, i.e. dissolve their citizenship.
[32] List and Pettit, *Group Agency*, 32.

organization does not deviate from fulfilling the prescribed criteria of agency. LP are particularly concerned about failures in group rationality.

1.5 The Discursive Dilemma

LP's worry about failures in group rationality can be explained by what they call "The Discursive Dilemma," a single phrase distilled from two well-known problems within the judgment aggregation literature (and elsewhere): one bearing the same name (The Discursive Dilemma), and one referred to as "The Doctrinal Paradox."[33]

The Discursive Dilemma highlights the potential for inconsistency in group decision making. The upshot of the Discursive Dilemma (drawn from Arrow's Theorem) is that there is no way to ensure aggregated group judgements will be consistent.[34] In the case of inconsistency, the group must have a way of choosing one or the other propositions. This means that some propositions will have to stand, not as a group judgement derived from individual judgment, but by some other method to be determined by the group. For example, the group may vote on three different but logically related propositions, which as a set, can be organized like an argument. In such a case, how should the group make decisions? If the vote on the "conclusion proposition" is what matters, why bother voting on the "premise propositions?" On the opposing view, why vote on the "conclusion proposition" at all when the group can instead vote on each "premise proposition" and let the conclusion follow logically? The two options correspond to what LP call the "conclusion based" versus the "premise based" decision procedures, respectively. Of course, this is more complicated than simply deciding on a case-by-case basis how to proceed. There is plenty of room for disagreement among members, especially in cases where deciding on a decision procedure will clearly change the group decision (and thus subject to decision procedure-hacking by individuals) - one's premise is another's conclusion. A long-lived group agent will have many past decisions to keep consistent with current decisions and many "arguments" that share propositions. LP's specific worry is about rational failures at the group level - a group agent should have its own

[33] Lewis Kornhauser and Lawrence Sager, "The one and the many: Adjudication in collegial courts," *California Law Review* 81, no. 1 (1993): 1.
[34] Franz Dietrich and Christian List, "Arrow's theorem in judgment aggregation," *Social Choice and Welfare* 29, no. 1 (2007): 19-33.

reasons for a decision, as it would not be rational to appeal to an inconsistent set of (individual) reasons as justification for the group agent's decision.

The Doctrinal Paradox arises in a legal context, specifically, in cases of determining liability. Suppose we have three judges voting on whether a defendant is liable or not. They may not be in agreement, in which case, a two-thirds majority can settle the matter. But unlike the Discursive Dilemma, the propositions supporting each judge's conclusion are the same. In other words, all are in agreement about which propositions are to count as reasons for the conclusion in each judge's individual reasoning. This means, that unlike in the Discursive Dilemma, there is agreement on which propositions are the premises and which is the conclusion. However, there may remain disagreement on the truth of the propositions. The Doctrinal Paradox then, is a problem about whether to use a premise-based or conclusion-based decision procedure.

The difference made by one decision procedure over another is far from trivial. In the case of the Doctrinal Paradox, it could mean the difference between a guilty or non-guilty verdict. With the Discursive Dilemma, it makes the difference between individual and group responsibility, and LP would argue, whether the group could even be a group agent at all. In a case where no individual or individuals can be held responsible, say because every individual action was minuscule, then the group may escape responsibility entirely, because we would have no basis (group reasons) to hold the group agent itself responsible (independent of its members).

LP use these examples to argue that, for the group to be considered a decision-making unit (e.g. a *judicial panel* as opposed to just three judges), members should "collectivize reason,"[35] making consistent past and present decisions, as well as ensuring the group has its own group-level reasons from which group conclusions follow. The way to do this is with a premise-based decision procedure,[36] in which the group considers *sets* of logically related propositions, ordering some as premises and others as conclusions. While individuals may hold different reasons for their own conclusions, a premise-based procedure makes clear and

[35] List and Pettit, *Group Agency*, 58.
[36] In actuality, LP prescribe a more deliberative process, e.g. a "straw vote" (62). Nonetheless, it is the clean distinction between the conclusion-based and premised-based methods that provide the relevant theoretical understanding.

explicit the logical decision-making processes of the group, and only those reasons considered as premises by the group will matter for (i.e., affect) the group decision. With a premise-based procedure, the members vote on premise-propositions but not the conclusion-proposition. The conclusion-proposition follows logically from the premises. The premise-based procedure, then, is a strategy to avoid inconsistencies, i.e., avoid the consequences of the Discursive Dilemma.

To summarize, the Discursive Dilemma shows that there is no voting procedure which can avoid inconsistent judgments. LP effectively demonstrate this with a few simple models (e.g., the three-judge panel determining liability) of highly puzzling cases in which the majority of individual judgements is at odds with the overall group judgment (what Frank Hindriks calls "disagreement cases").[37] It is in this sense, with nothing else mysterious intended, that one could say group agents have "minds of their own."[38] Thus, by illustrating cases of the Discursive Dilemma, LP demonstrate not only that groups can overcome these challenges and achieve rational unity, but also that groups can be autonomous agents in their own right.

1.5.1 Illustrations of the Discursive Dilemma

The Discursive Dilemma shows us that different methods of aggregating individual attitudes can affect the group attitude in surprising ways. For example, you could tally up individual judgments and then aggregate these into a group judgment, say by way of majority. However, if you were to tally up judgments on the *reasons* for those judgments, you might come away with a different group judgment.

To explain this difference in aggregating an overall decision from individual decisions, consider a case in which we might think of individuals' reasons as judgments about whether certain conditions have been met.

[37] Frank Hindriks, "How autonomous are collective agents? Corporate rights and normative individualism," *Erkenntnis* 79, no. 9 (2014): 1569.
[38] Philip Pettit, "Groups with minds of their own," In *Social epistemology, eds.* Alvin Goldman and Dennis Whitcomb (New York: Oxford University Press, 2011), 242.

	Condition 1	Condition 2	Final Judgement
Person 1	*Y*	*N*	**N**
Person 2	*Y*	*Y*	**Y**
Person 3	*N*	*Y*	**N**
Group Result:	*(Y)*	*(Y)*	$\rightarrow Y/\downarrow$N

Table 1.1: An illustration of the Discursive Dilemma

Suppose that it is the case that both Condition 1 and Condition 2 must be met for a positive final judgment (i.e. there being a suppressed premise or some rule stating that Conditions 1 and 2 jointly entail the conclusion). Each individual must judge for oneself whether the conditions have been met in order to make a final judgment. Were we simply to pool together the final judgments of each individual as indicated by the right most column in the table above, the group judgment would be "No" by majority (two "Nos," one "Yes" in the vertical column labelled "final judgment"). However, if we were to count the majority judgments on whether each condition was met (in italics), the group judgment would be "Yes." When adding "horizontally" we get a "Yes" result, but when adding "vertically" we get a "No" result. The "horizontal" ("Yes") group judgment adds previous group judgments together – the judgments added being the group judgments on the conditions themselves (in parentheses). The "vertical" ("No") group judgment adds individual final judgments (emboldened) together - judgments which are themselves based on whether each individual believed both conditions were met.

We might generalize such cases by thinking of them in terms of more standard arguments. In the example above the two premises are the two conditions and the conclusion their conjunction. Adding "horizontally" is what LP call a "premise-based" or "premise-centered" aggregation method, in which groups "collectivize reason."[39] This method of aggregation renders the group attitude as "Yes" despite the fact that a majority of group members' individual attitudes on the final judgment are "No." In other words, the group believes, *at odds with the majority of individuals in the group,* that the two conditions have been met.

[39] List and Pettit, *Group Agency*, 56; Pettit, "Groups with minds of their own," 250.

While counter-intuitive, it is this method that ensures collective rationality. Ensuring collective rationality is prioritized above "maximizing responsiveness to the views of individual members" (aggregating individual conclusions).[40] That is, some group level propositions must *not* be derived by aggregating individual judgments, but rather be determined by the group's view on other propositions.

The problem LP want to draw attention to is the fact that individuals within the group may make the same judgment for different reasons. In Table 1.1, Person 1 and Person 3 have the same final judgment ("No") but each has different reasons for that judgment. Person 1 judges that Condition 1 was met, but that Condition 2 was not. Person 3 judges that Condition 2 was met, but that Condition 1 was not. For a group agent to be rational (and thus be an agent on LP's account), it cannot abide contradictions or inconsistencies. For group agency, it matters how attitudes are aggregated. The premise-based method suggested by LP ensures that the group's final judgment follows logically from the group's judgement of the premises, not from the conclusions of individuals. By using this method, the group can be consistent when citing its reasons for its judgment (i.e. premises supporting the conclusion). The premise-based method also rules out, to some extent, the manifestation of individuals' hidden or ulterior motives, as the group agent's decision can only come about by way of what are clearly agreed upon as the relevant reasons (nominated propositions which the group agrees to vote on at the group level, i.e., the group-level "premises").[41] As we will see, this transparency plays an important role in LP's account of group responsibility.

The above example features a conclusion that follows from the conjunction of the two premises. Other examples feature disjunctive conclusions or conclusions that follow from a conditional premise.

Imagine a group of three environmental experts that constitute a group agent responsible for deciding whether or not to advise the government to introduce a carbon tax.

The motion (conclusion):

> The government should introduce a carbon tax. (Q)

Logically related propositions (premises):

[40] Pettit, "Groups with minds of their own," 251.
[41] "To some extent," is an important phrase. Some ongoing threats to group unity, and thus threats to group agency accounts, are documented in section 5.2.3.

There is a toxic level of carbon in the atmosphere. (P)

If there is a toxic level of carbon in the atmosphere, the government should introduce a carbon tax. (P→Q)

These related propositions should be organized like an argument in order to determine what the group should deliberate and vote on.

P (premise 1)

P→Q (premise 2)

Therefore, Q (conclusion)

LP argue that in order for the group agent to make a judgement about the motion, they should *not actually vote on the motion*, but rather vote on the logically related propositions (i.e., the premises). In the illustration below (Table 1.2), the majority vote of each premise will collectively determine the group output – a decision about whether to recommend the carbon tax. Based on a premise-based approach, the group agent judges the recommendation should be made.

	P	If P, then Q	Q
Expert 1	*Y*	*N*	**N**
Expert 2	*Y*	*Y*	**Y**
Expert 3	*N*	*Y*	**N**
Group Result:	*(Y)*	*(Y)*	→*Y*/↓N

Table 1.2: A second illustration of the Discursive Dilemma

Note that the group has a say on which premises they are considering as a group. Again, this helps to rule out ulterior motives that the conclusion-based aggregation method could admit.

Consider a situation in which Expert 1 would vote for a carbon tax for his own reasons, different from those of others.[42]

The motion: The government should introduce a carbon tax. (Q)

	P	[R]	P→Q	[R→Q]	Q
Expert 1	?	Y	?	Y	Y

R: One result of a carbon tax would be significantly less air pollution
R→Q: If a carbon tax would result in less air pollution, the government should institute a carbon tax

Table 1.3: A variation supplementing the second illustration of the Discursive Dilemma

Imagine a scenario in which Expert 1 endorsed the motion with other propositions featuring in his consideration. Suppose his endorsement of Q is completely unaffected by his views on P and P→Q.[43] That is, Expert 1 has reasoned individually. He has reasoned logically, with his endorsement of Q following from R and R→Q. His reasons support his conclusion. Similarly, we could imagine a case in which all the experts endorse Q, but for different reasons. In these situations, *the group* does not have *its own* reasons to support its conclusion, only a mass of individual reasons, which at the group level, spells inconsistency. No *group reasons* could be cited in support of a group decision to endorse Q.

Suppose the panel of experts vote using the conclusion-based method (the method which does not ensure group-level rationality). Suppose the result is unanimous: everyone voted "Yes" on proposition Q. Later, when called to account for the *panel's* decision, you attempt to provide group justification retrospectively. You could go to each member and ask for their reasons supporting Q. You ask Expert 1, who cites both R and R→Q as his reasons. You write down his answers then present them to Expert 2. "How about you, Expert 2? Did you also come to conclude Q based on R and R→Q?" She responds that her view on both R and R→Q is "No," and furthermore, that her view on propositions R and R→Q did not feature in

[42] In this particular case, Expert 1 votes for respectable impartial reasons, thus "ulterior motives" need not imply something indecent or corrupt, rather just that Expert 1's reasons are different official reasons considered by the group.
[43] If we consult Expert1 about all four premises, maybe we would discover that Expert 1's endorsement of a carbon tax would be overdetermined (i.e. he endorses both sets of reasons). Or maybe Expert 1 would reject P→Q, but accept Q because of R and R→Q. To make the present point, we need not know his opinion on P and P→Q - we just need to know they played no role in coming to his decision about Q.

her judgement regarding Q (e.g., she instead reached her judgement by considering P and P→Q). In such a scenario, your attempt to justify group judgments based on individual judgements has already failed (even before consulting Expert 3). Part of your explanation would end up something like this: "The panel has judged that Q due to R and not R."[44] Trying to provide group reasons where there are none results in inconsistency. The only explanation that makes sense would be one offered at the level of individuals (experts, not *the panel*).

The premise-based aggregation method solves this problem. Deciding on which propositions to vote on is the decision of the group, as is deciding on which to count as premises and conclusions (i.e. "sequential priority procedure").[45] Therefore, Expert 1 can cast his vote on P and P→Q, but not R or R→Q. The group only recognizes propositions P, Q, and P→Q. Therefore, unlike Expert 1's individual reasoning process in the previous example, propositions R and R→Q do not factor into the group's decision, nor even Expert 1's decision because he is not asked for a decision on Q at all.[46] Experts are not asked to vote on Q, so any reasons particular to their view on Q do not factor in the group's judgement on Q (unless those reasons are shared and voted on collectively). Voting according to differing individual reasons, ulterior motives, and suppressed premises is even less likely if P and P →Q are voted on at different times over the group agent's life, in isolation from the "argument" that renders the group's view on Q. In such cases the decision about Q is already made by looking to the group's historical votes on logically related propositions (P and P →Q) - though of course, the group can revote on those propositions (or rearrange their logical sequence). According to LP, such "adjustments" may be called for from time to time in order to avoid inconsistency, and thus, maintain group level rationality. For example, a new motion is favored but its endorsement would be inconsistent with past votes; the group must decide against the motion, revise their endorsement of the other proposition with which it is at odds, or alter the "sequence" of the propositions such that the inconsistency dissolves.

[44] This may seem at odds with my comment in the previous footnote. Expert 2's view on R was only revealed when trying to gather the reasons in support of the motion. *But that's the point* – her view on R shouldn't be relevant, or it should, and if relevant, everyone in the group should have voted on R (as I will go onto make clear).
[45] List and Pettit, *Group Agency,* 56.
[46] Exceptions include cases where individuals act on their "outcome-oriented preferences," defecting from their epistemic duties *qua* group member, e.g. "disingenuous voting" (111).

That latter point is significant, as recursive voting (or "voting") is what makes reasoning and reflection possible for group agents. The above illustrations are rigid examples, and LP themselves suggest real group deliberation will likely be less rigidly constructed, decision procedures more likely to be "inexplicit" (hence the scare quotes on "voting"). For example, a group agent should be mindful to note potential inconsistencies, as doing so helps them decide on the sequence on which to consider logically related propositions. Therefore, often a "straw vote" (or "straw poll") would be more appropriate than a one-off voting system.[47] The group can produce judgements, then return to consider where to adjust or possibly compromise, not only to preserve rationality (e.g. avoid inconsistency), but also to properly reason and reflect on what is being considered. The upshot is that nothing is ever decided that cannot be undecided so long as issues are properly attended to at the group level. All things can be revised.

Illustrations of the discursive dilemma show not only that group rationality can be maintained, but that groups are autonomous agents in their own right, or so LP argue. Explanations of group judgements are not so straightforwardly explained by the sum of individual judgements. Indeed, the aggregation function employed by group agents is holistic, as it operates on sets of propositions rather than individual propositions. Mark Jensen makes LP's point most clearly: "Group attitudes and actions are tied to individual attitudes and actions, but at the level of sets, not at the level of propositions."[48] This is the "supervenience thesis:" everything the group does is done by individuals, to be sure, but group judgements supervene holistically on individual judgements.[49] There is no straightforward one-to-one type mapping of group properties to individual properties. According to LP, this non-reducibility factor is what makes group agents not only rational agents, but autonomous agents.

1.6 Ideal Group Agents

Complex group agents are clearly more sophisticated than minimal agents, especially large groups such as corporations or states that organize themselves by way of an intricate division of labor and hierarchies of authority and expertise. For LP, it will be these further capacities

[47] List and Pettit, *Group Agency*, 62.
[48] Mark Jensen, " Is our group an agent? Do we want it to be?" *Journal of Moral Philosophy* 12, 4 (2015): 548.
[49] List and Pettit, *Group Agency*, 66.

that we should demand of groups if they are to be recognized as agents in society. We must note, however, that within LP's account, these further capacities are expressed as *desiderata* for an ideal group agent, not as requirements that must be met by every group agent.[50] For example, drawing on empirical support, LP argue that groups will benefit from democratization, decomposition, and decentralization. Democratization involves harnessing the wisdom of the crowd by instituting a majoritarian democratic structure.[51] Decomposition and decentralization allows groups to divvy up decision tasks to sub-groups (e.g. based on expertise), and consistent with a premise-based procedure, have those sub-groups deliberate and render a judgement on premises, from which a group judgement (i.e. conclusion) will follow.[52] (The focus of decomposition is on divvying up the issues; the focus of decentralization is on organizing the sub-groups of individuals to which issues are assigned).

Another aspiration of the ideal group agent is the achievement of what LP call "incentive-compatibility" and "control."[53] The control d*esideratum* expresses LP's concern about the power and potential dominance of group agents over their members. On their view, an ideal group agent should be organized so as to carefully preserve the rights and freedoms of its members. Incentive-compatibility refers to the familiar challenge of aligning individual and group interests. LP have established that a group can be a rational agent with beliefs and goals of its own. But of course, individual members are themselves rational agents, pursuing their own goals, which may or may not be in tension with group goals. Therefore, an ideal group agent should be designed "not for maximally well-behaved or idealized group members but for real people, who behave strategically when this is expedient."[54] As one would expect, LP make no grand proclamations about how to solve this timeless dilemma, but they do make distinct two separate strategies. The "organizational route" describes a path to incentive-compatibility in which you take the preferences, beliefs, and behavior of individual agents as given, and adjust the group structure to suit. The "behavioral route" does

[50] "Ideal group agents" is not a category used by LP. I make use of it here to remind the reader that these further capacities achieved by a group agent are surplus to the requirements of agency. One potential point of criticism is that LP may do too well in convincing us of the importance of these further ideal qualities such that maybe these qualities should be listed among the minimal conditions, and not just the further "desiderata." Or alternatively, as I suggest in chapter 5, we could benefit from a range of nuanced categories that capture kinds of agents or even groups that only approximate group agents.
[51] List and Pettit, *Group Agency*, 90.
[52]List and Pettit, *Group Agency*, 92-97.
[53] List and Pettit, *Group Agency*, 104, 129.
[54] List and Pettit, *Group Agency*, 104.

the opposite, trying to change individual preferences, beliefs, and behavior.[55] Nothing concrete is offered here, but they do suggest that one way of taking the behavioral route is through the achievement of solidarity (though that is not a term they use). Referencing Bacharach,[56] they speak of the group adopting a "we-frame," in which individuals identify themselves with the group (e.g., language in the plural first-person).[57] LP also highlight the potential to harness the fact that people care what others think about them, hinting at a possible achievement of unity through cooperation and reputation.

A reader gets the impression that throughout their book, LP implicitly lay out what can be seen as an ontogeny of group agents, that is, the development of sophisticated group agents through various states of development (a journey from minimal to sophisticated agency). We can note not only different sorts of agents, but the schematic of a certain progression: from minimal agents, to rational-autonomous agents, and as we will see in the next section, even to *personal* agents. Of course, some firms, for example, will be founded with principles and procedures such they qualify as fully rational agents from day one. LP are not implying that groups always come to be agents by way of development or "growth" (although their thoughts on corporate responsibility do suggest some will need to be "parented").[58] Rather, as noted briefly in section 1.2, they set minimal requirements of agency, but then introduce constraints and requirements indicative of higher standards they see as desirable for group agents. If all of these *desiderata* are met, a group agent can even be considered a *personal agent.* (Worries about what is and should be required versus merely desirable for group agents will be addressed in sections 2.2.3 and 2.3, and partially motivates section 5.1).

1.7 Responsibility and Personhood

For many, the primary reason for wanting to count groups as agents is to be able to hold them responsible. We want to be able to hold group agents responsible so as to protect individuals, for example, from the harmful actions (or inactions) of powerful groups. On the other hand, groups themselves may desire agent status so as to gain certain rights and privileges (e.g. to

[55] List and Pettit, *Group Agency*, 124.
[56] Michael Bacharach, *Beyond individual choice: teams and frames in game theory*, (Princeton: Princeton University Press, 2018).
[57] List and Pettit, *Group Agency*, 128.
[58] See section 2.3 on the "developmental rationale."

enter into contracts).[59] LP are particularly worried about cases in which the blame cannot be pinned on any particular individual(s). For example, there may be cases where each individual makes tiny, individually insignificant contributions to the group action (such that no individual can bear (enough) responsibility), or cases in which the cause cannot be traced any further down than the group as a whole. In such cases, not being able to hold groups themselves responsible means no agent is held responsible. This result creates what we intuitively feel to be an unjustified responsibility deficit with nowhere for the blame to land.[60]

On LP's account, group agents are fit to be held responsible when they:

> 1. are able to recognize when the group agent faces a normative choice between a right and wrong action.
> 2. are able to gather and understand what counts as evidence or reasons for one choice over another.
> 3. have the "control required for choosing between the options."[61]

Consistent with my description of a "group agent ontogeny," LP even allow that groups may be held responsible in an instructive type of way. This is thought to be similar to how we hold children to standards they could not possibly yet meet, so as to shape their character such that they will act responsibly as adults, rising to the expectation of first, their parents, and second, society. This also has the ring of "making an example out of" irresponsible group agents, so that observant group agents will form or if already formed, better their procedures to avoid undesirable consequences (i.e. liability). This topic will be further explored in section 2.3.

LP argue that group agents can reason, make normative judgements, and be held responsible for their actions and are therefore persons.[62] They cite Locke, Hobbes, Pope Innocent IV, and the western legal tradition more generally (going back centuries) as consistent with their view.[63]

---

[59] In section 5.1 I will suggest a taxonomy in which moral patients (e.g. infants) have rights but no obligations and agents (e.g. group agents) have obligations but no rights.
[60] In a more specific discussion we would likely want to keep the notion of "responsibility" and "blame" cleanly distinct. For our purposes in this thesis, the distinction does not matter.
[61] List and Pettit, *Group Agency*, 158. This amounts to arguing that groups have control in the same way we assume individuals have control (not making any claims about something like "corporate freewill").
[62] List and Pettit, *Group Agency*, 176.
[63] E.g., Thomas Hobbes, *Leviathan: With Selected Variants from the Latin Edition of 1668*, ed. E.M. Curley (Indianapolis: Hackett Publishing, 1994), 166.

LP take their view of personhood to be analogous to their view of mind. What qualifies as "mental" should not be based on whatever it is that physically makes up the mind. To say something is mental is to say it plays a certain role or discharges certain functions.[64] Similarly, "the mark of personhood is the ability to play a certain role, to perform in a certain way."[65] To perform as a person means to "perform effectively in the space of obligations." More specifically, it is "to be party to a system of accepted conventions, such as a system of law, under which one contracts obligations to others, and… derives entitlements from the reciprocal obligations of others."[66] LP argue groups are capable of performing in this way by virtue of their capacity to reason and self-regulate, as laid out in LP's *desiderata*. Basically, group agents are able to intentionally, and continually check for, detect, and correct errors in rational and normative processing.[67] Emphasized here, however, extrapolated from the technical capacities required for reasoning, is the personal agent's ability to deploy those capacities "to function within the space of mutually recognized obligations,"[68] relating to other agents "as sources and targets of addressive claims."[69] In a recent paper, Pettit emphasizes that group agents can self-represent in their reasoning and interaction with other agents, and that individuals engage with them via a *conversational stance* (implying discursive and evaluative abilities beyond those of a mere rational agent, the kind "projected" by taking up the intentional stance), relying on group agents to communicate consistently and competently with the communicative capacity and unified voice like that of a human agent.[70]

So then, group agents are able to self-represent, reflect, reason, and perform competently in the normative realm. However, despite making a case for the possibility of group persons, LP still feel that group persons should be subject to greater scrutiny and not be entitled to equal standing with individual persons.[71] They cite the asymmetry of power between individuals and groups and their commitment to normative individualism as reasons for this. Yet it remains unclear whether these reasons are meant to serve as any real justification for this position. Relevant challenges and criticisms to this position will be developed in chapter 5.

---

[64] List and Pettit, *Group Agency*, 170-171.
[65] List and Pettit, *Group Agency*, 171.
[66] List and Pettit, *Group Agency*, 173. "Persons are agents that are capable of having demands made upon them, and of making reciprocal demands in turn, within a system of norms," (170).
[67] List and Pettit, *Group Agency*, 178.
[68] List and Pettit, *Group Agency*, 177.
[69] List and Pettit, *Group Agency*, 176.
[70] Pettit, "Group Agents Are Not Expressive, Pragmatic or Theoretical Fictions," 1656.
[71] List and Pettit, *Group Agency*, 18.

To summarize, LP argue that groups can meet not only the minimal conditions of agency, but further criteria, in light of which they are fit to be held responsible. Such group agents are then accountable to, and members of, the moral community.

Chapter 2
Agency and Rational Unity

In this chapter I aim to provide some relevant additional context to LP's work and the group agency literature more generally, offering a mix of descriptive and critical content, organized by topic. In Section 1, I will highlight the work of Miller and French as important precursors to LP's work. In Section 2, I explore the concept of agency more generally featuring work by Korsgaard and Rovane, providing contrasting views with an emphasis on rational unity and practical rationality – concepts I will further develop in relation to biological individuality in Chapter 4. In Section 3, I will consider what it means to become an agent and develop as an agent, drawing on aspects of the previous section, paving the way for an argument in Chapter 5 calling for more work on nuancing importantly distinct categories of "agent." In Section 4, I feature recent work by Collins on various types of collectives. Only some of these collectives qualify as group agents - others might need more nuanced categories to satisfy a key *desideratum* that motivates LP and others: plugging the responsibility gaps that sometimes arise from collective actions. In Chapter 5, I will feature Collins' categories in a model inspired by the literature on biological individuals and populations.

2.1 Foundations and Motivations – Miller and French

2.1.1 The Insufficiency of Joint Action Accounts

What primarily interests some is simply explaining collective actions whether or not the individual actors are integrated in such a way as to meet the criteria for group agency. Do those who act together *share* intentions? Or do individual intentions come together to somehow form a group-level intention?

Seumas Miller argues that joint actions can be explained without appeal to group intentions.[1] Joint action is when individuals act together to bring about a goal. The goal they work to achieve is *the same goal* - one realized by a single state of affairs.[2] To explain, consider two strangers, Bob and Billie. Both may share the goal of retiring at age 65, but their working and

[1] Seumas Miller, *Social action: A teleological account* (Cambridge University Press, 2001).
[2] Tollefsen, *Groups as Agents*, 31.

saving does not contribute to the same end. Billie can achieve her goal while Bob fails to achieve his. Bob and Billie only share the goal of retiring at 65 in the way they might share the quality of having brown eyes.[3] Miller explains individuals' intentions to work together in terms of shared knowledge and interdependent beliefs about others doing their part to reach the collective goal.[4] Jill's effort to lift the left side of the table only makes sense if, and is only exerted on the understanding that, John lifts the right side.

One might object to Miller's parsimonious proposal. We might think that an action is, by definition, the kind of thing that is guided by an intention, in which case either a group level action needs a group level intention, or it is not an action. Bratman would respond that intentions are like plans or promises.[5] They not only guide, but also constrain an agent's actions in accordance with them (constraints not provided by shared goals alone). Would we not need group level intentions to play a similar role in the case of joint actions?

Another problem is that requiring shared knowledge that others work for the same end does not seem plausible for large groups like big corporations. Yet corporations act. So then, shared knowledge is neither sufficient nor necessary for group action.

Unlike LP, Miller does not take groups to have intentions or goals. Rather individuals can share intentions and goals such that they work together to achieve that goal by way of a joint action. In such a view, attributing intentions to groups would be a mistake. However, if we are sympathetic to LP's case for (some) groups with a mind of their own (e.g. as illustrated by the premise-based decision procedure in instances of the Discursive Dilemma), we would conclude Miller's explanations are insufficient in some cases (cases in which an account of group *agency* is arguably needed).

One of LP's main motivations is to address what is often referred to as responsibility gaps or deficits.[6] In a case where no individual is responsible, no agent is responsible. If all group actions are only the actions of individuals (and not a group agent), Miller has to bite the

[3] What they share is the *having of the same kind of goal*.
[4] Tollefsen, *Groups as Agents*, 29.
[5] Michael Bratman, "Précis of shared agency: A planning theory of acting together," *Journal of Social Ontology* 1, no. 1 (2015): 1-5.
[6] E.g. Stephanie Collins, "Collective responsibility gaps." *Journal of Business Ethics* 154, no. 4 (2019): 943-954; John Hasnas, "The Phantom Menace of the Responsibility Deficit," In *The Moral Responsibility of Firms*, eds. Eric Orts and N. Craig Smith (Oxford: Oxford University Press, 2017): 89-105.

bullet in regards to moral responsibility gaps. Powerful groups can perform harmful acts and, where no individual(s) can be held responsible, there will be no agent to hold responsible.

2.1.2 List and Pettit's Predecessor

Accounts of joint *action* like those of Miller can be seen as what LP are *reacting to.* Previous work on group *agency* like that of Peter French can be seen as what LP are *building upon*.[7]

French's work on corporations as persons (later revised to "moral actors") was an important precursor to LP's account. The two accounts share a focus on decision-making.

Hobbes claimed that a group could co-opt the unity of a spokesperson and count as an agent.[8] Groups can be their own spokesperson (e.g. committee), and thus groups could be agents even without nominating a *single* human to funnel their words and actions. French, however, dispenses with the need for a spokesperson to serve as a unifying mechanism or vehicle of vicarious agency (individual, committee, or otherwise). Instead, unity is achieved in a corporation by way of its internal decision structure. The Corporate Internal Decision (CID) structure "accomplishes a subordination and synthesis of the intentions and acts of various biological persons into a corporate decision… it *incorporates* acts of biological persons."[9]

Corporate decision-making is based on its structure, with some people answerable to others, some taking up this task, others taking up that task. Individual duties are laid out by a corporation's rules of conduct, charter, bylaws, etc. As corporations are formed for a purpose (e.g. to sell widgets), there will be a basic policy (i.e. corporate principles, rules of conduct) on which the corporation functions.[10] When operating consistently with corporate policy "it is proper to describe it as having been done for corporate reasons, as having been caused by a

---

[7] Not only French, but others as well, some in dialogue with French at this time through a series of papers mostly in the context business ethics and responsibility. Manuel G. Velasquez, "Why corporations are not morally responsible for anything they do," *Business & Professional Ethics Journal* 2, no. 3 (1983): 1-18; Larry May, *Sharing responsibility* (Chicago: University of Chicago Press, 1992); Rita Manning, "Corporate responsibility and corporate personhood," *Journal of Business Ethics* 3, no. 1 (1984): 77-84; Kenneth Goodpaster and John B. Matthews, "Can a corporation have a conscience," *Harvard business review* 60, no. 1 (1982): 132-141.

[8] David Copp, "Hobbes on artificial persons and collective actions," *The Philosophical Review* 89, no. 4 (1980): 579-606.

[9] Peter A. French, "The corporation as a moral person," *American Philosophical Quarterly* 16, no. 3 (1979): 212. Emphasis added.

[10] Peter F. Drucker, *Concept of the Corporation* (New York: John Day, 1946).

corporate desire coupled with a corporate belief and so, in other words, as corporate intentional."[11] The CID provides a way to unify corporate thought and action and works to maintain this unity simply by virtue of its structure. The division of corporate roles, rules, and duties, (the "organizational chart") provides checks to weed out inappropriate personal influence on corporate activities; these interactions "dilute" personal reasons that might make their way into reports and recommendations (e.g. French suggests academic peer review is just such a diluting process).[12]

We can note French's CID and LP's premise-based straw-pole decision procedure both play similar roles. Generally, both French and LP are citing the presence or lack of certain organizational structures as determinative of which groups are agents.

French calls collections "aggregates."[13] Being collections, aggregates cannot survive changes in membership. If one person is removed from an aggregate, the resulting aggregate is not the same aggregate. More relevant is that they are not agents. We can point to aggregates as being causally responsible for something, but not morally responsible (in the sense of being blameworthy or accountable for their collective act). "Moral responsibility predicates cannot be legitimately ascribed to aggregate collectivities" because "aggregates simply fail the tests for membership in the moral community."[14] (Mapping types of responsibility to kinds of collectives will be explored in more detail in section 2.4).

French offers the phrase "conglomerate collectivity" to contrast with aggregates in that conglomerates can survive changes in membership: "an organization of individuals such that its identity is not exhausted by the conjunction of the identities of the persons in the organization." Conglomerates have decision procedures and organizational structures with an enforced code of conduct and a tight integration of defined roles and responsibilities carried out by its members. In other words, conglomerates are group agents. French articulates many of the important points about non-reducibility that LP's later account illustrates using the Discursive Dilemma. That a conglomerate holds some view or is responsible for some

[11] French, *The Corporation as a Moral Person*, 213.
[12] An organizational chart maps "the interdependent and dependent relationships, line and staff, that are involved in determinations of corporate decisions and actions" (213).
[13] Note that LP make a distinction between mere collections and groups, whereas French is happy to call both kinds (e.g. sets unrelated people, all those in the world that share a physiological features) "aggregate collectivities."
[14] Peter French, *Collective and corporate responsibility* (New York: Columbia University Press, 1984), 11.

decision does not entail that any given individual within the conglomerate holds that view or can justifiably be held responsible for the decision.[15]

Raymond Pfeiffer has criticized French for not being careful enough with his definition of conglomerates.[16] Pfeiffer argues that French defines conglomerates in regard to their non-reducibility and persistence through membership changeover, but then lists separately characteristics by which we might identify conglomerates as distinct from aggregates (e.g. decision procedure, defined roles, etc.). Since it is by this latter list that we can know which groups are intentional agents, French should have been more careful to include these "characteristics" within a proper definition (i.e. as necessary and sufficient conditions). Extending a bit of charity (French's definition and list of characteristics are introduced without otherwise obvious signposting or fanfare within two paragraphs of each other), Pfeiffer's point may be profitably valued not as a devastating account-crunching zinger leveled at French, but more as an important commentary on what we will need in a clear definition, and indeed, a herald of the clearer articulation we get from LP's account.

Other related criticisms are levelled at French based on the distinction between aggregates and conglomerates – the former a mere collection, the latter a group with an organized structure and decision procedure – as being too clean, with no room for "blurry" cases of collectives which may fall somewhere between a mob and a corporation.[17] If what French claimed to be a difference of kind is truly a difference of degree, his account likely faces a number of problem cases.

We should note that perhaps a form of this criticism could also be levelled at LP. On the one hand, LP's account makes very clear the categorical differences between these types of groups. But on the other hand, they lay out a spectrum from minimal to personal agents and allow for inexplicit procedures and non-intentional group formation. Having acknowledged that, I still think the kinds of cases Pfeiffer claims to be problematic should not be seen as problematic just because they (purportedly) show there exists a continuum between mere collections and group agents (i.e. his criticism questioning the clear demarcation of

[15] French, *Collective and corporate responsibility,* 14.
[16] Raymond S. Pfeiffer, "The central distinction in the theory of corporate moral personhood," *J Bus Ethics* 9, no. 6 (1990): 473–480.
[17] E.g., Robert Ware, "Review: Collective and Corporate Responsibility," *The Philosophical Review* Vol. 96, no. 1 (Jan. 1987): 117-119.

agglomerates and conglomerates, e.g. his neighbourhood scenario, in which neighbours take inexplicit cues acting in informal roles without any intentional organization or explicit hierarchy). Rather, in my view, Pfeiffer's critique stretches the notions of unintentionally formed groups and inexplicit procedures too far such that it is hard to view the groups used in the examples as real group agents, in LP's sense.[18] Whether or not Pfeiffer's criticisms are straw-manning French, they would be attacking a strawman if applied to LP given LP's clearer articulation of group agency criteria and attention to group level rationality. This is an instance in which we can observe that portions of LP's work can be seen as building on French's, completing and supplementing the same kind of group agency account.

However, concerns like those of Pfeiffer do offer a stress-test for both the scope and elasticity of criteria in LP's account. It is important to remember that, while exploring whether groups could be agents is philosophically interesting in itself, an account of group agency is often offered as a *means* to meet a concern that motivates philosophical work – usually moral responsibility. Not all groups we will want to hold responsible will be agents, and thus an account of group agency will not be an effective means. In section 5.1.5, I will suggest a possible pathway to exploring the "messier" terrain when considering a spectrum of more or less agents, possibly unintentionally formed or with an inexplicit decision procedure.

Lastly, French argues that corporations are persons, not just agents. LP likewise say group agents can be group persons. In section 5.1.3, I will argue that this is a mistake and offer a preferable taxonomy that more carefully distinguishes those with and without interests that matter. So, while LP's work can be seen as building on, and patching the gaps of French's account (e.g. with a clearer focus on rationality requirements demonstrating non-redundancy), it in other ways repeats the same mistakes.[19]

2.2 Agency and the Unity of Agency

To further appreciate the relevant issues with group agency it will be useful to explore the concept of agency more generally. As LP's emphasis is on rational unity, we might pay

[18] Incidentally though, independent from criticism of French, the existence of what we may take to be in-between collectives (e.g. group agents to a degree) organized through inexplicit decision procedures, are potential motivators for what I offer at the end of section 5.1.
[19] Indeed, in later work, French himself corrected this particular mistake, when he revised "moral person" to "moral actor." Peter French, *Corporate Ethics*, (New York: Wadsworth, 1994).

special attention to the unity of agency. In chapter 4, I aim to build a conceptual bridge between the unity of purpose exhibited by biological individuals and the rational unity exhibited by agents. The accounts featured in this section round out LP's more technical treatment of group-level rationality with more of an emphasis on the practical and normative aspects of rationality. Importantly, from these accounts, we can see that even conceptions of agency intended for application primarily towards individuals render plausible the notion of group agency. The attention given here to rational unity and the agent-constitutive notion of a "rational point of view" lays the groundwork for portions of chapters 4 and 5.

2.2.1 Normative Agency, Rational Unity, and Identity

Now I will turn to the work of Christine Korsgaard, introducing it by way of J. David Velleman. Velleman highlights an agent's capacity to make "future directed decisions."[20] But ironically this diachronic aspect of autonomy in the present could be seen as compromising one's autonomy in the future. Were one to decide now what part of a project one should undertake at some future time, it seems as if one's future self is like a mindless robot, following an algorithm in order to execute the former self's plan. So how do agents make and carry out decisions about the future without compromising future agency? The answer, Velleman argues, is that an agent provides reasons for one's future self to cooperate.

Agents make decisions that their

> future selves will be determined to execute of their own volition; and the only way to determine our future selves to do something of their own volition is by giving them reason to do it. Unless we can commit ourselves today in a way that will generate reasons for us to act tomorrow, we shall have to regard our day-older selves either as beyond the control of today's decisions or as passive instruments of them.[21]

From a synchronic perspective it is as if we are talking about a community in which one produces reasons to gain others' support for one's view. Reason allows our present selves to

[20] David Vellemen, "Deciding how to decide," In *Ethics and practical reason*, eds. Cullity, Garrett Cullity and Berys Gaut (Oxford: Oxford University Press, 1997): 46.
[21] Velleman, "Deciding How to Decide," 46.

produce what we think will be convincing reasons for our future selves to carry out our plans in the same way that one produces reasons to convince others of their view in an argument.[22]

It would seem clear then, that for rational unity to obtain (e.g. means-ends reasoning), one's diachronic agency must be intact in terms of agent-stages that are in agreement (i.e. consistent, rationally unified). There seems then a plausible analogy to draw between group agents and individual agents. Individual agents face the challenge of diachronic agreement similarly to how group agents face the challenges of synchronic agreement. But of course, we can consider both types of agency as facing both types of problems (diachronic and synchronic integration). Christine Korsgaard suggests a more direct, synchronic analogy between individuals and groups:

> When a group of human beings occupy the same territory… they have an imperative need to form a unified state. And when a group of psychological functions occupy the same human body, they have an even more imperative need to become a unified person.[23]

The relevant message from Velleman and Korsgaard here is about *practical necessity* in relating what it means to be an agent and to be the same agent.

Different philosophers have emphasized different aspects of agency over others. At times what is emphasized is being a subject - having a capacity to experience with an interest that those experiences are not unpleasant or unsatisfactory. Other times our focus is on being an actor, considering what one should do. (In chapter 5 I will introduce a taxonomy in an attempt to better distinguish these two aspects and establish their logical independence). Korsgaard cites Hume and Parfit as among those that take a person to be primarily a "locus of experience." This "passive" aspect of agency can be contrasted with the "active" aspect of agency. Korsgaard frames the former as a "utilitarian starting point" and the latter as

---

[22] This squares with a well-regarded evolutionary psychology argument that reasons are for social consumption, as well as the psychological literature from which we have learned that we do not appear to have any special "internal" access to much of the contents of our minds, e.g. what we believe and why we believe it. See Hugo Mercier and Dan Sperber, *The enigma of reason* (Harvard University Press, 2017).

[23] Christine Korsgaard, "Personal identity and the unity of agency: A Kantian response to Parfit," *Philosophy & Public Affairs* 18, no. 2 (1989): 115.

"Kantian."[24] For Korsgaard, agency is *achieved* by conformity to norms and therefore she warns that agency should not be reduced "to a mere form of experience."[25] For Korsgaard, the focus of agency is personal identity, not experience, for it is the former that grounds the unity of agency. Korsgaard argues that "the conception of ourselves as agents is fundamental to the standpoint of practical reason."[26]

Whereas a reductionist like Parfit may conclude that the purported unifier of experiences should be less of a concern than the quality of the experiences themselves, Korsgaard conceives of agents as separate individuals with clear boundaries delimiting not merely experiencers, but actors (i.e. deliberators, choosers).

Korsgaard's main point here, as informed by Kant's view, is that even if reductionism suggests there is no metaphysical reality to being a concrete subject (whatever that might mean), there nonetheless exists a practical necessity for taking oneself to be a real, unified, self-same agent over time. Korsgaard concludes that agency is the right concept with which to explore personal identity, not experience.

Korsgaard's account of agency is a normative one that is compatible with reductionism (because she holds no assumptions or metaphysical givens). Drawing on Kant, she notes two ways we can view ourselves: as objects of theoretical study (concerned with predictions and explanations) and as agents (concerned with choice and justification). With the former, agents are no different than the natural stuff of the world - forces and events that play out all around us. But as the latter, we view ourselves as centers of action and deliberation. It is in this way that we must view ourselves "when we occupy the standpoint of practical reason."[27] Regardless of what metaphysical reality may or may not underlie our subjectivity, the

[24] According to Korsgaard's normative account of agency, "we constitute our own agency or activity by adopting certain laws to govern our actions." Christine Korsgaard, "The Normative Constitution of Agency," in *Rational and Social Agency: The Philosophy of Michael Bratman*, eds. Manuel Vargas and Gideon Yaffe (Oxford: Oxford University Press, 2014), 196. See also Christine Korsgaard, *Self-constitution: Agency, identity, and integrity* (Oxford: Oxford University Press, 2009).
[25] Korsgaard, *Personal Identity*, 131-2.
[26] Korsgaard, *Personal Identity*, 132. Even without regard to agency, anyone who has engaged in a philosophical exploration of personal identity must confront the fact that the idea of personal identity itself is difficult to make sense of, yet we cannot meaningfully live without it, making it seem a simultaneously urgent yet futile topic (similar to what many will feel regarding the subject of freewill).
[27] Korsgaard, *Personal Identity*, 120.

standpoint we occupy (i.e., that of an agent) "is forced upon us" anyway, as we must make choices, navigate the world, and coordinate our thoughts and activities.[28]

The naturalistic concept of agency (e.g. Donald Davidson[29]) states that **"**an agent is active when her movements are caused or causally guided by her own mental states or representations."[30] What constitutes an action, as opposed to a mere behavior or event, is that it follows a certain causal pathway flowing from the subject's intentional states, e.g. a belief-desire pair, forming an intention, leading to a motion.

Korsgaard sees the naturalistic concept in which actions are caused by mental states as unsatisfying due to a familiar regress: "the mental state itself presumably has prior causes, so it is not clear why its operation should especially represent the agent's own activity or spontaneity, any more than anything else."[31]

By contrast, the normative concept implicates the actor's identity such that others' responses to the actor's actions are responses *to the actor.* We understand one's actions "to reflect something really essential about her, to represent her in some way." An action is "unlike other events whose causes in some way *run through* an agent,"[32] but rather is meant to be "a movement, or the effecting of a change, that is backed by the agent *as a whole*."[33] In practice, we take the actions of others as "legitimate or at least appropriate grounds for responses like love and hate, liking and disliking, gratitude and resentment." Such responses are directed at the *whole* agent. "Unity… is necessary to ground the special kind of attribution that is distinctive of agency: attribution to the agent himself." [34]

In the latter quote, we observe one of Korsgaard's main points: agency is a practical necessity, not a metaphysical given. Regardless of whether one is in some sense a real subject with identity preserved over time, there nonetheless exists practical reasons for treating oneself as such. These practical reasons include enabling the basic means-ends reasoning, in

[28] Korsgaard, *Personal Identity*, 113, 119, 120.
[29] Donald Davidson, *Essays on Actions and Events* (Oxford: Oxford University Press, 1980).
[30] Korsgaard, *The Normative Constitution of Agency,* 294.
[31] Korsgaard, *The Normative Constitution of Agency,* 194.
[32] Korsgaard, *The Normative Constitution of Agency,* 193. Emphasis added.
[33] Korsgaard, *The Normative Constitution of Agency,* 193.
Emphasis added.
[34] Korsgaard, *The Normative Constitution of Agency,* 194.

line with Velleman's concern with agent-stages being in rational agreement. There is a "practical necessity imposed upon you by the nature of the deliberative standpoint… We both presuppose and construct a continuity of identity and of agency."[35]

Korsgaard sees "unity implicit in the *standpoint* from which you deliberate and choose" (a point Carol Rovane often emphasizes in her own work, as we will see in the next section).[36] Korsgaard suggests there is a "raw necessity of eliminating conflict among your various motives" due to the fact that one must act, and one has "only one body with which to act."[37] Korsgaard shows little interest in making a distinction between a rational point of view and a phenomenal point of view (though as we will see, Rovane does). However, I think Korsgaard is simply highlighting the practical grounds on which one considers oneself a single, unified agent. I do not think, however, that the distinction is lost on Korsgaard, though she does articulate the interesting position that the unity of consciousness "is simply another instance of the unity of agency;"[38] consciousness is a "lack of perceived difficulty in the coordination of psychic functions."[39]

But we might also worry that Korsgaard's comment about "one body" either rules out group agents, or does not actually apply to all agents (i.e. only to single-bodied agents). Korsgaard has at the least, not ruled out the idea of group *action*: "On the view of action as normatively constituted, unification is an achievement, not a given… There is nothing that prevents people from acting together in the most literal sense of the word."[40] The careful reader will note it sounds like Korsgaard is talking only about collective *action* and not necessarily the possibility of a collective *agent*. Perhaps as a Kantian, she is wary of the latter. To return to Velleman's point at the beginning of this section, having a single body with which to act forces my future, past, and present psychological episodes to act in unity, i.e. as *an* agent. Because group agents don't possess this constraint of physical embodiment, a group agent is not forced to act or conceive of itself as an agent in the same way as an individual human. This is why a decision procedure is crucial to LP's account (and arguably, the plausibility of group agents more generally), and perhaps why, lack of incentive compatibility is an ever-

[35] Korsgaard, *Personal Identity*, 113.
[36] Korsgaard, *Personal Identity*, 111.
[37] Korsgaard, *Personal Identity*, 111.
[38] Korsgaard, *Personal Identity*, 119.
[39] Korsgaard, *Personal Identity*, 119.
[40] Korsgaard, *The Normative Constitution of Agency,* 40.

present threat to group-level unity (these potential threats will be developed further in section 5.2.3 with other aspects suggested for future work in section 6.2).

We have surveyed Korsgaard's view of agency as a practical necessity – the unifying standpoint of rational deliberation. The notion of group agency challenges our assumptions about the pairings of agents with bodies. LP meet this challenge by arguing that a group can take up this standpoint and possess a rational point of view all its own (over and above those of its members). In the following section, we will push these ideas a bit further, arguably, too far.

2.2.2 Dissociative States, Fragmentation, Disunity

Carol Rovane holds what I will call an "exclusive view" of agency, which does not permit of agents nested within other agents. Rovane suggests what may seem a radical reframing of what it means to be an agent. Before getting into Rovane's account however, it might be best to first consider in more detail the possibility of dissociative states and similar cases in which a single human may host many agents (i.e. "alters"). I will therefore set the stage for Rovane by featuring portions of Kennett and Matthews' work on dissociative disorders.

Kennett and Matthews are interested in psychological conditions that threaten agency.[41] Personhood, they say, requires the ability to

1. undertake and complete projects,
2. participate in social relationships, and
3. occupy social roles in the service of a variety of institutional ends.

To develop the capacities necessary for personhood requires unified agency. Dissociative conditions threaten this unity.

Drawing on Korsgaard and Velleman, Kennett and Matthews hold that autonomous agency requires "competent deliberation" and "the capacity for self-control over time."[42]

[41] Jeanette Kennett and Steve Matthews, "The unity and disunity of agency," *Philosophy, Psychiatry, & Psychology* 10, no. 4 (2003): 305-312.
[42] Kennett and Matthews, "The Unity of Disunity of Agency," 308.

Regarding ongoing projects and long-term activities, such as earning a degree or cultivating a relationship, Kennett and Matthews claim that you cannot simply swap out one person for another and still say that the new person is continuing the same project. "The very identity of such projects as these depends on the relation they bear to the particular individual who created them in the first place."[43] We might think that because group agents can survive changes in individual membership that in such cases where individuals are replaced, group projects do remain the same projects. Indeed they would, but this comparison is misleading. The appropriate comparison would be to swap out the agent (i.e. the group agent), not its component parts (individuals).

Kennett and Matthews illustrate the disunity of one with dissociative identity disorder, describing Mary, a mother who cannot keep tabs on her children nor guarantee she will fulfil a promise. If Mary shares her body with other resident agents (alters), then presumably Mary is absent (or partially so) when it is not "her turn." But is this really disunity of agency? Could Mary, as just one of many of her personalities, also not be a narcoleptic under this description? It would seem what Mary experiences is a disruption of consciousness. This may be a useful comparison to keep in mind when considering group agents who are only "activated" during e.g., board meetings, and otherwise remain "dormant." If there is disunity of agency, is it about the dormancy itself – about the fact that one's temporal orientation is incongruent with others in the moral community (i.e. an interrupted, intermittent life span)? Or is it less direct – that such an incongruity leads to the effect that one cannot satisfy social norms (e.g. keep promises)?

Regardless, there is an important sense in which unity is lacking when we think of Mary and her alters as a single-bodied community. In this way, the group agent, Mary and her alters, does experience a disunity of agency, but Mary does not. To achieve unity Mary's reasons would have to also be reasons *for her alters,* much as "normal" agents do over time, as suggested by Velleman.

That she suffers from DID is not sufficient for claiming Mary lacks unity. However, upon further seeing that her various personalities are uncooperative and rationally inconsistent at

[43] Kennett and Matthews, "The Unity of Disunity of Agency," 308.

the "group" level, we can then justifiably claim disunity.[44] In principle, one with DID could deliberate "as a group" and overcome the challenges of DID by achieving agency as some groups do. In other words, it is conceivable that one could have DID with many different personalities, each expressing their own beliefs and desires, and yet all come to form an agent with a decision procedure (however unconventional) such that rationality is achieved at the whole person level. It seems then that exceptional cases in which agency is achieved despite DID would fare better than one with narcolepsy, whose intentions can be thwarted randomly (i.e. without a means of functioning while dormant and without control over when one is dormant).

This way of thinking is of interest when exported from Kennett and Matthew's particular argument that DID should be treated as a disorder of agency. Similar to Velleman, Kennett and Matthews have us consider agency in terms of a threshold of psychological integration. In some sense, all paradigm personal agents (i.e. adult human beings) are functionally integrated groups. In chapter 4, I will emphasize the constitutive role of agential unity not only as common currency between group and individual agency, but between both and biological individuality.

Now I will move on to feature Rovane's normative account of agency. For Rovane, an agent is not one who merely fulfills normative requirements of rationality, but rather one who is *committed* to meeting them. Contrast this with LP's account, in which an agent would require something more than rationality (recursive intentional processes - namely what they call "reason") to demonstrate such a commitment. Rovane's account is more traditional, in that possessing true rationality is evidenced by this commitment, and thus by rationality itself (by definition), regardless of any instances of failures in rationality. Rovane's minimal agent necessarily (to count as an agent) acts to correct mistakes, whereas LP's minimal agent could not assess whether or not it was acting in accordance with its rational "programming." This aim or commitment to meeting rational standards is a more demanding requirement for agency.

For Rovane, an agent deliberates, makes judgments, and chooses between options in

[44] Even discussing this case is difficult because it is not clear when and whether the name "Mary" refers to a group of personalities or a single personality.

accordance with the demands of rationality. An agent is one that surveys its own attitudes and makes "all things considered" judgments (consisting of all the agent's own attitudes, but no others' attitudes).[45] The phrase "its own" emphasizes the practical sense of "rationality" in addition to the sense denoting the requirements of consistency and coherence (i.e. logical or theoretical rationality). This aspect is important for Rovane, holding that a necessary condition of agenthood is deliberating from a unified point of view.

> Agency is always exercised from a rational point of view. For there can be no such thing as agency unless agents have reasons on which to act, and these reasons comprise an agent's rational point of view… a deliberative point of view from which the agent can determine what would be best to do, all things considered, in the broadest sense of 'do' that comprehends all intentional activities, including mental activities.[46]

Furthermore, Rovane's characterization of a rational point of view is "essentially normative."

> A rational point of view is something in which contradictions and conflicts *ought* to be resolved, implications *ought* to be accepted, preferences *ought* to be ranked, means and consequences *ought* to be taken into account – and more generally, it is something from which all-things-considered judgments *ought* to be reached and implemented.[47]

From here, we can ask what *kinds of agents* can possess *a* rational point of view. Rovane makes no assumptions about a one-to-one pairing of agents with human beings. She often refers to humans as "sites" - bits of space-time that may host a single agent, multiple agents or part of an agent. The intentional activity of an agent may spread over multiple human beings or reside within only part of a human being. The bounds of what is considered in the agent's "all things considered" judgments may be considered the true boundaries of the

[45] Carol Rovane, "What is an Agent?" *Synthese* 140, no. 1/2 (2004): 183.
[46] Carol Rovane, *The bounds of agency* (Princeton: Princeton University Press, 1998), 85-6.
[47] Rovane, *The bounds of agency*, 21

agent.[48]

Rovane makes a radical claim: “Human beings can continue to make their contributions to a group effort only at the cost of their rational unity as individuals”[49] Despite this claim that humans can neither be part of group agents nor spawn multiple sub-human agents without fracturing their own agency, she does not claim that doing so violates norms of rationality.

> Insofar as such things are worth doing, they may provide human beings with reasons to integrate into group agents or fragment into multiple agents. And, because this is so, it is misguided to suppose that group and multiple agency would necessarily involve transgression against the normative requirements of rationality.

Rovane writes that agents are

> obliged to deliberate from their own points of view about what it would be best to think and do, all things considered. But, if I’m right, their deliberations should include the effort to identify potential agents that they might help to bring into existence through fragmentation or integration. Their deliberations should also include an evaluation of the relative worth of what such smaller and larger agents could accomplish, as compared with what these deliberating agents can do as the size agents they presently are. The result may well be a decision on the part of an agent to end its life through fragmentation or integration. If the decision is justified, then, at the moment of decision the agent does achieve the sort of rational unity within itself that it is the proper aim of all agents to achieve. And yet, the implementation of this decision will serve to undermine that very unity. That is the sense in which human beings do not stand under a rational obligation to achieve unity within themselves. They can reason their way out of existence, by deciding to fragment into multiple agents or to integrate into group agents – or, as I think it usually comes to pass – they can decide to go in for a combination of both.[50]

On that note, I will take up a critical response to some aspects of Rovane’s account. What is

[48] Carol Rovane, "Is group agency a social phenomenon?" *Synthese* 196, no. 12 (2019): 4873.
[49] Rovane, *What is an agent,* 189.
[50] Rovane, *What is an agent,* 197.

related in the passage above seems paradoxical, reminiscent of similar problems encountered in the philosophy of personal identity. The literature on human enhancement, for example, poses problems of "uploading" one's consciousness to a computer or augmenting oneself physically such that we cannot be sure that doing so would leave the *same* person intact; in other words, the person deciding may not be the person who benefits.[51] More relatable, L.A. Paul writes about deciding to become a parent or graduate student, each being a "transformative experience" in which the one choosing could not possibly, from their limited vantage, evaluate the consequences of the outcome, which may be conferred upon a fundamentally *different person* (the transformed person).[52]

How can one be said to act rationally when making the choice to cease existing? Even when we are careful to keep distinct the notion of rationality as self-interest from the notion of rationality as deliberation from a particular point of view, Rovane's claim is still extreme: does involvement in a group agent really require me to sacrifice my own agency? Rovane might say that the value of the goal (e.g. made possible by a group agent one incorporates themselves into) may be seen as more valuable than one's own life as an individual agent. But this level of sacrifice seems on par with dying for one's country, for one's ideal, or less abstractly, one's loved ones. Is involvement in a group agent really so drastic as this?

Within the final phrase of the quotation above is a clue: "as I think it usually comes to pass." This talk of creating new agents and putting to death existing agents sounds dramatic unless we are careful to play by Rovane's rules. Agents are not whole human beings to begin with. Perhaps Rovane's claims only seem implausible when we import our own definitions and intuitions (e.g. matching one-to-one brains to agents). In other words, perhaps we should read Rovane as re-describing something that regularly happens. She is not prescribing some radical departure from ordinary human life, as a literal reading might suggest. Maybe the more relevant criticism here is that Rovane's real subject is about *roles, projects,* and *personas*, and not *agents* as we might view agency more traditionally. Rovane says most group agents "absorb only some of the energies of their human constituents, leaving behind another agent of significant, but not fully human-sized, proportions."[53] So perhaps the real

[51] E.g., Nicholas Agar, "On the irrationality of mind-uploading: a reply to Neil Levy," *AI & society* 27, no. 4 (2012): 431-436.
[52] Laurie Ann Paul, *Transformative experience* (Oxford: Oxford University Press, 2014).
[53] Rovane, *What is an agent,* 194.

value of Rovane's work, is showing how we not only spend our time, but divide up our projects according to different rational points of view (or if not actually different rational points of view, different considerations depending on the various roles we fill).

Taking her at her word, Rovane would have us believe that for individual and group processes to run parallel, they must be the processes of two distinct agents (i.e. my less-than-human-sized personal agent, and the bit of me that, with others, constitutes the group agent). Consider a CPU meter as referenced on a personal computer. This is bit of software that monitors and reports to users the amount of processing power being deployed (i.e. the use of processors, amount of RAM). Such a meter, among other things, reveals the number of "background" processes taking place at any given moment on one's computer. A computer's resources are not 100% dedicated to each task at any given time. By analogy, it seems plausible that while I am in a committee meeting (i.e. a deliberation process of a group agent), I might also be daydreaming, remembering a personal experience, mulling over my to-do list, or reconsidering my own *personal* political views as informed by the group discussion. Maybe Rovane makes an interesting point, that like computational resources measured in percentages, perhaps something like 19% of my attention is on my inner musings, and 81% is on the group task at hand.

But to carve these percentages of usage into separately existing agents will, to a critic of Rovane's account, seem a further, unnecessary, and unparsimonious claim. This is especially true if, like computational resources, the attention we give to various activities is fluid, with boundaries negotiated second to second. Intuitively, the CPU analogy is superior to conceptualizations of discrete agents of different sizes, especially if the latter entails new and different agents popping in and out of existence with changing boundaries.[54] So if the deployment of resources is a better fit, one skeptical of Rovane's view might think this provides further evidence that the very interesting aspects of identity Rovane is concerned with, while important, are better explored in terms of projects and roles and commitments of a single agent.[55]

---

[54] A phenomenon that pseudo group agents could hack, harness, and exploit for members' gain. See "agent-when" groups, described in section 5.2.3.

[55] The CPU analogy raises a point about epistemic access. It will be instructive to note a particular asymmetry between an individual member and the group agent of which that member is a part. I, as a faculty member, can recall intentional episodes of a group agent. But a group agent cannot recall episodes of me. A university faculty cannot recall thoughts in my life, but I can recall "thoughts" in the faculty's life. You might reply, not exactly, a member is just remembering group activities from that member's individual point of view. But if that's the case,

One can make a case to be made then, for redeploying Rovane's argument as a *reductio* against the very view of agency she offers. Rovane writes, "in cases of group agency, the rational unity that is achieved at the level of the group typically precludes rational unity at the level of its human constituents within their whole lives, though it can be realized within parts of those human lives."[56] Critics will find it peculiar that Rovane stands behind this notion of an agent. However, thinking through her account is useful. If read as a *reductio*, we might conclude that the rational-point-of-view account is not sufficient to capture what we care about regarding agents (as opposed to roles and project-contributors). If we take on this view of agency, we end up with counter-intuitive boundaries to agents. "No further metaphysical condition, beyond the existence of intentional episodes standing in the right relations, suffices for the existence of an individual agent with its own deliberative point of view."[57] Rovane's view leaves us with a world full of large and small non-overlapping agents born of fission and fusion, and can thus serve as a foil for those offering an inclusive account (such as LP). Individual humans cannot preserve their entire rational unity and at the same time partially constitute a rationally unified group agent. For those more sympathetic to an inclusive view, Rovane's account, while interesting, is not suitable to accommodate the topic of group agency.

On this *reductio* reading then, we could maintain that granting agency solely on the basis of deliberating from a rational point of view *is* a problem if there are appreciable differences between group and individual agents. LP claim there are such differences and are committed to normative individualism (privileging the interests of individual agents above group agents). In section 5.1, LP's commitment to normative individualism will earn its own criticism, owing somewhat to the ambiguity of the word "individual." While LP can escape the bizarre consequences of a view like Rovane's because of their commitment to privilege sentient human beings as importantly distinct from other agents who possess a rational point of view but no sentience (e.g. group agents), their own account of what makes an agent an

---

the individual's memory is testimony to the individual's *point of view* being prominent during that time! In reply, one could invoke a distinction between one's rational point of view and one's phenomenal point of view. But even so, a relevant asymmetry remains. The phenomenal experiences undergone in one's capacity as a member may later be drawn upon as an individual, seamlessly from their point of view, in some sense of "point of view" that is not as easily disentangled as the theoretical distinction would have us believe. Anyhow, this asymmetry might be reason to resist an exclusive account of agency (i.e. one that does not permit of nested agents).

[56] Carol Rovane, "Group agency and individualism," *Erkenntnis* 79, no. 9 (2014): 1663.

[57] Rovane, "Is group agency a social phenomenon?" 1478.

agent (and specifically, what makes a person a person) is not sufficient to distinguish their account from one like Rovane's.

So the foregoing critique of Rovane motivates concerns for LP that I will address later in this thesis. There is another important take away from Rovane's account. I have been suggesting Rovane's conception of agency is not suitable for humans and human groups on an inclusive view of agency. An ontology of agents and parts of agents independent of the kinds of entities they comprise leaves out a lot of what is important e.g., to our moral considerations. But in the case of our theorizing about biological individuality, I will argue the converse. By focusing on what is and is not a biological individual, we can leave out an important part of the picture, that while a given entity's individuality is in question, we can (and often do anyway!) in the meantime describe the relevant phenomena in terms of agency.[58] The usage of agential heuristics in our theorizing is not limited by what we take to be *real* agents and *real* individuals (in whatever sense). Even parts of wholes are only *contingently* parts. Some of the work presented in this section emphasizes agency as something achieved and maintained. I aim to emphasize the same about biological individuals (in which paradigm individuality is the achievement of groups, and other kinds of individuals are defined in terms of ongoing cooperation or collaboration). For example, cooperating somatic cells in a multicellular organism are always potential defectors, in the sense that a mutation event is an instance of fragmentation (I will further develop this sort of example in section 5.2). Saying too much along these lines here would be premature; but before concluding this section, I will say a little more for the sake of prefiguring later portions of the book.

For Rovane, rational fragmentation multiplies the number of rational agents, however differently "sized." For Rovane, fragmentation of a human life simply means there is no commitment to make the entirety of that human life rationally unified.[59] Fragmentation occurs when "not all of the thoughts and actions associated with their brains and bodies proceed from the same point of view."[60] Think of what this means for members of group agents. If the intentional activities of my brain and body are carried out from the group agent's point of view, "they are literally *not mine*."[61]

[58] It may seem curious how we justify the leap between talk of deliberation and rationality to biological individuals and their traits. This will be the focus of chapter 4.
[59] Rovane, "Is group agency a social phenomenon?" 4880.
[60] Rovane, "Group agency and individualism" 1665.
[61] Rovane, "Group agency and individualism" 1667.

I argue that this perspective may be useful in some cases of biological individuality. The *behavior* (or some other trait) observed in some biological critter at our focal level may in fact be the *action* of a biological entity at another level. When the actions of an individual curiously benefit the group at cost to themselves, we might infer the action belongs to the group but was performed through the individual. What seemed irrational at the individual level was rational at the group level. When the actions of a group curiously benefit one member at cost to itself, we might infer the action belongs to the individual but was performed through the group. Similar stories can be told in a range of biological contexts (e.g. meiotic drivers, suicidal behaviors in those hosting parasites, male sterility in plants).

So then, Rovane's preparedness to liberally attribute agency, while viewed by some as unpalatable when applied in its own context, primes us to think differently in another context: that of biological individuality. I will develop these ideas further in section 5.2.

2.2.3 Critiquing "Minimal Agency"- Self Consciousness and the Capacity to Fail

We have discussed that LP allow that a minimal agent may be rational without knowing it is rational or without being committed to rationality. They suggest as an example a robot programmed to set blocks upright: it has motivational states (that the block should be upright) and representational states (observing that the block is currently not upright) and the capacity to act on those states (change the block's position such that the representation now matches the motivation).[62] Some will say that LP's simple robot which is rational by virtue of its programming cannot be an agent, minimal or otherwise. Why? Because rational unity is not an achievement for the robot; it is not a project undertaken and sustained by the robot. Perhaps the ever-present *threat of disunity* is required for agency. Korsgaard suggests something like this, having in mind self-conscious human beings. I say "something like" because she addresses it the other way around: "The complexity introduced into the human psyche by self-consciousness introduces a standing threat of disunity that must be overcome before the agent can act."[63] So, for any self-conscious agent, the threat of unity is inherent, and this motivates her normative account. Perhaps Korsgaard would allow that LP's robot

[62] List and Pettit, *Group Agency*, 19.
[63] Korsgaard, *The Normative Constitution of Agency,* 193.

could be an agent in some very different sense, but as her Kantian account of agency is geared around human beings, it is more likely she would say the robot does not share that fundamental challenge to overcome and so is just not an agent. Reflexivity ("reason" for LP) involves a fundamental partitioning - what is it to be self-conscious other than to stand apart from and analyze oneself? For those like Korsgaard and Rovane (and others like Frankfurt whose account emphasizes higher-order reasoning) who take agency to always be a product of effort and will (something forged or achieved), disunity must be a possibility - a possibility that would be realized in one's failure to achieve unity (i.e. to integrate various parts, whether separate functions, desires, or different times in one's life).

LP often speak of the necessity of "robustness" – the assurance of consistent performance across a range of circumstances and inputs to the group. However, we might at least question whether "robust" is the right concept (or at least examine the perceived need for robustness). Should we really be in the business of insisting group agents be set up to be safe from failures? Should we not rather ensure they are capable of recognizing and acting to correct failures? Built-in robustness seems inferior to corrigibility with a capacity to learn from failings. One could argue that LP's prescription amounts to paternalistic programming such that instead of ensuring agency, such rigid constitutive criteria keep groups from ever realizing agency.

This criticism is really two criticisms:
1. As Korsgaard might argue, if there is no risk of failing to achieve agency, then one is not an agent.
2. If one cannot know whether one is rational or not, one is not an agent.

The second has already been addressed and will remain controversial. For LP, this further capacity is "reason," which need not be possessed by minimal agents. Though, despite their permissive criteria for minimal agency, LP are also interested in more demanding conceptions of agency (e.g. agents in a more demanding sense that can be morally responsible).

The first criticism is related to the second but does not demand self-awareness in regard to rationality. Rather it highlights the idea that rational agents are sometimes irrational – and that irrationality can become a threat to agency itself. We usually don't take instances of

irrationality to dissolve one's agency such that they are no longer agents. Nor do we assume e.g. human agents must build up for themselves ways to ensure they are never irrational. What we want of developing agents is to attain something like "anti-fragility,"[64] such that they can learn from instances in which the unity of agency was threatened, momentarily toppled, or otherwise unclear (in other words, we want instances of irrationality to shore up and increase rationality as opposed to the expectation that it leads to further degeneration of rational capacities). Of course, rather than instances of irrationality, one can fail to achieve agency all together. But firstly, in the case of humans, these people are usually considered unwell and are therefore special cases; secondly, as I will suggest in section 2.3, it is mysterious to speak of the "one" that fails to achieve rational unity (a reductionist would say an aggregate of psychological episodes failed to integrate into an agent). In chapter 6, I will outline for future work a parallel to this bootstrapping problem in evolutionary biology in which we see individuality as the evolved character of a collective which must exhibit sufficient qualities *as a unit* on which selection can act!

Whatever the case, there is need for some stronger response to these criticisms. If LP stand by the notion of minimal agency, then the criteria seem insufficient, or at least incompatible with many accounts of standard (non-group) agency. However, if what LP really want of agents are those further capacities built "on top of" minimal agency, then they need to consider raising the bar for minimal agency such that it reflects what they really feel is important and essential for agents.[65]

To put perhaps too fine a point on the potential incommensurability between accounts like those of LP and Rovane, note that what Rovane calls a "reductionist account" of agency stipulates that the rational requirements of agency are "irreducibly normative."[66] Holding and abiding by a commitment to be rational is not the same as one organized, disposed, or programmed to ensure rational processes. This irreducibility, she argues, is evident in the fact that an agent may fail to live up to such requirements. While the labels sound similar, Rovane's "reductive" account and LP's "minimal" agency could not be more different. Whereas LP are happy to say that minimal agency is achieved when rational requirements are

[64] Nassim Nicholas Taleb, *Antifragile: Things That Gain from Disorder* (New York: Random House, 2012).
[65] One might respond that we use the term "agent" in different ways. The latter point motivates a more definite and precise classification, which partially motivates section 5.1
[66] Rovane, "Is group agency a social phenomenon?" 4872.

met (e.g. when consistency and transitivity obtain), Rovane will say that no agent is an agent accidentally or due to programming - agency is grounded in commitment to those norms, whether they are achieved or not.

2.2.4 Summary

To summarize this section, the normative aspects of rational agency allow us to observe meaningful parallels between the agency of individual and group agents. Korsgaard and Rovane's emphasis on the practical rationality they take to be constitutive of agency are a useful supplement to LP's focus on theoretical rationality (e.g. consistency of group judgments). I believe that this survey of Korsgaard and Rovane also shows that the topics of "agency" and "group agency" are much the same topic in regard to particular aspects – the notion of "rational point of view" operates the same in both cases. This has considerable implications for any account that is blind to the other aspects that would otherwise differentiate between groups and individuals. Important differences between groups and individuals, and what these differences mean for LP's account, will be discussed in the first half of chapter 5.

In regard to "minimal" or other nuanced categories of different kinds of agency (or "agency"), we might find a way both to represent some of these importantly different categories on a continuum (e.g. in terms of ranked levels of agency or achievement or degrees of paradigm agency), while preserving other non-continuous categories that we hold matter intrinsically. I will suggest ways we might pursue this project in section 5.1, drawing on useful concepts from the individuality debate within the philosophy of biology to build on and supplement LP's account.

In chapters 4 and 5, I will argue that the language of agency is of great use when giving the best explanations about biological individuality, analogizing between the unity of intentional states and biological traits. The notion of a rational point of view can be applied to the "as if" agency of biological entities that meet sufficient conditions of unity. (It is equally important to note that when such unity fails, a rational point of view might instead (or also) be attributed to the "fragmented" component, i.e., treating it as an agent, not merely as part of an agent).

2.3 Agential Growth and Development

The previous sections – in particular Korsgaard's account – brings us to the question of whether agency is an achievement. Puzzles about what we might call "agential ontogeny" can be roughly categorized into questions about 1) creation or origin, and 2) growth or development. Let us attend to these in turn.

Who or what is the achiever of agency? Who or what forges the agent? What pre-agent undertakes agency as a project to achieve? Or for Korsgaard's Kantian view, who or what adopts a maxim to make a law, such that living in accordance with the law makes of one an agent? The bootstrapping nature of Korsgaard's (and others', e.g. Rovane's) normative conception points to an important difference between individual agents and group agents. Group agents acquire rational unity when individual rational agents take up the project of achieving and maintaining rationality at the group level. The forgers, the achievers, are individuals who form a group. Once achieved, we *may* say it was the group who achieved it. We can say this because the achievement was achieved *by means of rational agency* (i.e. the rational agency of individuals). So it would actually be *better* to say that the group's continued agency is an achievement of the group, but its initial agency was the achievement of individuals. The real initial achievement is at the level of the individual.

It is not clear that there is a meaningful analogue in the case of an individual. To do so risks positing sub-personal homunculi, and perhaps homunculi all the way down.[67] Whereas individuals achieve initial group agency, it is not clear what plays the role in achieving initial individual agency.

Here is another way of asking the question. If norms are something we can fail to abide by, if unity is something one can fail to achieve, who or what fails? How can the attempt and the achievement be part of the *agent's* story? Perhaps the effort to maintain agency is credited to the agent, but how can the birth or forging of an agent be credited to that agent?

[67] If homuncular functionalism is to be endorsed as a hierarchy of complex tasks or structures that can be boiled down to simpler and simpler tasks or structures, then even so there will be a cut-off point in which this nested structure is not *agents* all the way down (so long as we do not equivocate about what counts as an agent).

Arguably, in most cases, individual human agency is a gift from the wider society of agents in which we are raised.[68] Norms are imposed externally and we rise to them, hammered into the semblance of our shapers. My agency is someone else's achievement – at least initially. I find that when considering Korsgaard's Kantian view, it is hard to rid oneself of the notion of some sort of abstract solipsistic state of nature where individual proto-agents were born into a world in which they must bootstrap themselves into existence as a subject distinct from their environment. Call this "The Bootstrapping Thesis," to serve as our foil. But if we acknowledge that one's agency is the product of others (e.g. our parents), an analogy between individual level and group level unity might be more plausible, as in both cases, the agent forged is not mysteriously responsible for forging itself when such forging seems to require agency to begin with. Taken to the extreme this is the antithesis of Korsgaard and Rovane's claim that agency is always that agent's achievement. Agents may (and perhaps must?) take on the maintenance of their own agency once agency is established, but as far as origins are concerned, one's agency is the achievement of another. Agents make agents. Call the latter view the "Collective Achievement Thesis."

Suppose we think the societies (or tribes, or families) that shape the norms of developing humans are themselves group agents. On this view, while group agents depend on individual agents for their existence, individual agents depend on group agents for theirs. Agents come from other agents - individuals from groups, and groups from individuals. I find this a fascinating thesis (that agent formation is both hierarchical and symmetrical) that I will sketch for future work in section 6.2.3.[69]

What remains unsatisfying however, is an endless regress: who were the first agents such that the intergenerational legacy persists? Here is perhaps where looking to evolutionary biology will prove useful. The unity of agency (or its precursors) may have arisen due to selection pressures. The social norms we rise to meet likely had their precursors such that a step-wise evolutionary story could easily be told. Korsgaard often draws upon the unity of mind and consciousness to make her point, emphasizing the challenge of unifying ourselves so as to coordinate activities with our solitary body. While the unity of agency is not a given, what is

[68] Ann E. Cudd, "Individual Agency as Collective Achievement," in *Proceedings of the XXIII World Congress of Philosophy 44* (2018) 5-9.
[69] There is also the possible case of a group agent creating another group agent, e.g., fraternities forming new chapters, companies starting franchises, etc.

given is sufficient for achieving it. While we cannot credit evolution with the project of unity we must all undertake in our lives (and as I suggest, *early* in our lives, and it being *our community* undertaking the project of *our* agency), we can credit evolution with capacities that can be broken down into selected traits, stretching back millennia, as temporally analogous to the explanatory processes of homuncular functionalism.[70] And in this, the endless regress is not so unsatisfying after all. Rather we have separated historical, evolutionary causes of agency from the within-lifetime cause of agency within a particular individual. Should we be unsatisfied with our ignorance of the earliest stages of evolution, then so be it - this problem is not unique to explorations of agency. Many in other fields, such as early life researchers and roboticists, are equally puzzled about the first of the first (e.g. how possibly could the first replicating molecules with metabolic capacities have arisen?). So if there is still an unsatisfying foundation here, it is just as much a problem for similar regresses in other domains (e.g. the origin of culture). In other words, how the first autonomous systems bootstrapped themselves into existence is one of our biggest questions. Generally, we do not tend to let the biggest, most foundational mysteries halt our more immediate investigations.

We should provisionally conclude two things:

1. The "Collective Achievement Thesis" has important implications. If individual agency is seen as a collective achievement, we can make more plausible analogies with group agency, as group agency is also a collective achievement. To forge an agent, however constituted, requires collaboration among agents that share the norms they will bestow upon, and reinforce in, the proto-agent.

2. We can distinguish questions about how organisms evolved the capacity for agency from questions about how agency comes about in individuals. We can put it this way: the bootstrapping problem for agency is as easily solved as would be the bootstrapping problem for reading. We would never consider the latter a problem, and that's the point. The mystery of reading is not a special evolutionary problem – our ability to read is easily explained by the mechanisms that make reading possible. The evolution of the mechanisms themselves is a

[70] See "homuncular functionalism" in Simon Blackburn, *The Oxford dictionary of philosophy* (Oxford University Press, 2005).

different story. We are the products of evolution such that the capacities we had were selected, or at least were not culled.[71] As advanced life forms, we are well past basic autonomy and proficiency at extracting resources from our environment - expertise that entails seeing ourselves as apart from our environment. We possess a suite of plastic traits that allow for a great deal of within-life adaptivity (e.g. learning). No one is born with the ability to read, but once literate, one can transform oneself into something extraordinary.

That agents are created or forged presents us with an apparent chicken-egg puzzle. In chapter 6, I will sketch for future work the hierarchically symmetrical version of the Collective Achievement Thesis I mentioned briefly above (that individual agents come about by way of group agents, and group agents come about by way of individual agents) by exploring connections to the evolution of biological individuality and related ontogenetic processes in which new individuals are produced or emerge. The evolution of individuality shares a similar chicken-egg puzzle to the picture of agents achieving their own agency in a bootstrapping manner in that the *preconditions* for selection are often thought to be the *products* of selection. Individuality at a higher level is an evolutionary achievement of groups of individuals at a lower level. It likewise takes individuals to comprise groups that can then be selected as a group that achieves sufficient integrated organization. Further, this can all be related to LP's notion of incentive compatibility, and the higher-level organization achievable by biological individuals with aligned evolutionary interests (see section 6.2.1).

Moving on, what about how agents develop? It is intuitive that agents do not spring fully formed out of the ether but rather bloom or grow (e.g. on the understanding that an infant is not an agent, but an adult is). LP and others with similar accounts often, implicitly or explicitly, have something to say also about the development or growth of agents.

Bjornsson and Hess first lay out the difference between agents and moral agents. Agents "must be capable of acting on the basis of their own beliefs, desires, and intentions," while moral agents are, in addition, "subject to moral obligations and accountable for their actions." Moral agents must be capable of "recognizing and acting on moral considerations, including respect owed to others." Bjornsson and Hess also add that a moral agent must be "capable of

[71] I put it this way trusting the reader knows I am not making specific positive claims that amount to adaptationist just-so stories (and certainly not about literacy, which is too recent a trait to have been selected).

acting freely in some relevant sense."[72]

Bjornsson and Hess then smuggle in another category, "fully-fledged moral agents," with the further criteria that these agents "must be capable of certain reactive attitudes, in particular those of guilt and indignation." This means that for a group to be a "fully-fledged" moral agent they must collectively be capable of reactive attitudes such as guilt – attitudes held at the group level. That is, the group itself must bear this attitude, not just (or even!) its members.[73] Bjornsson and Hess conclude that groups that are agents are capable of "states sufficiently similar" to guilt and indignation and so therefore also count as moral agents (in this paper Bjornsson and Hess argue that if arguments for group agency are successful, then the fact that they can have reactive attitudes follows).

Misselhorn approaches agency similarly, acknowledging levels of agency that arise with additional features or capacities: basic agents with autonomy and intelligence, goal-directed agents, intentional-agents, agents with higher-order intentionality.[74] Precedents for these sorts of hierarchies can be found in the early cybernetic literature.[75]

Mark Jensen likewise articulates his reading of LP in terms of stages or a nested structure, going from "core agency" (corresponding with LP's "minimal agency"), to "rational agency," to "autonomous agency," and finally "personhood."[76] While all that is required for agency on LP's account is minimal agency, successive "stages" can be seen as charting a trajectory towards the ideal agent. This is worth unpacking. A minimal agent has intentional states and the capacity to act on them. When this agent conforms to the standards of rationality (i.e. attitude to fact, attitude to action, and attitude to attitude) it is a rational agent. When the agent's actions and attitudes cannot be readily reduced, or easily explained in terms of its components (in the case of a group, its members, e.g. explaining group judgments from

[72] Gunnar Björnsson, and Kendy Hess, "Corporate Crocodile Tears? On the Reactive Attitudes of Corporations," *Philosophy and Phenomenological Research* 94, no. 2 (2017): 273.
[73] Bjornsson and Hess, "Corporate Crocodile Tears," 283.
[74]Catrin Misselhorn, "Collective agency and cooperation in natural and artificial systems," in *Collective agency and cooperation in natural and artificial systems*, ed. Catrin Misselhorn (London: Springer, 2015), 3-24.
[75] E.g. Arturo Rosenblueth et al., "Behavior, purpose and teleology," *Philosophy of science* 10, no. 1 (1943): 18-24.
[76] Mark Jensen, "Is our group an agent? Do we want it to be?" *Journal of Moral Philosophy* 12, no. 4 (2015): 542.

individual judgments), then the agent is an autonomous agent.[77] Lastly, if this agent operates fluently and competently in the space of social obligations, it is a personal agent.

Jensen highlights two of LP's *desiderata* for group agents: incentive compatibility and individual control (see section 1.6). Incentive compatibility means that individual goals and group goals are aligned sufficiently as to avoid individual-level and group-level conflict (e.g. group members acting to fulfill personal goals at odds with the group's goals). Individual control means that the group does not simply dominate its members. Jensen conceptualizes these two *desiderata* as side-constraints, but not as requirements of any of the aforementioned "stages" or "levels." The reason for this, I think, is that Jensen sees the side-constraints as applying specifically to group agents (as we would have to speculate a bit as to how such hierarchical conflicts exist literally or in some analogous way, for the individual and the individual's sub-personal processes).[78]

I will (in section 5.1) raise the criticism that LP blur the lines between what is required and what is ideal, and how requirements may (or should?) differ based on the kind of agent (individual or group). It might be that for a more generalized account of agency, we may want to recast as a requirement what would otherwise be considered aspirational (a point we considered at the conclusion of the previous section, 2.2). While the picture of "developing" agents may pave the way towards a set of different *desiderata* (e.g. a spectrum of agents and groups approximating agency, see section 5.1.4), it introduces ambiguity: when is a group a *real* agent? Should side-constraints be requirements?

In chapter 5, I will suggest a need for more nuanced concepts, perhaps a plurality of agents not unlike the plurality of individuals on offer from pluralists within the biological individuality debate. It is possible to lay out different categories based on the type of agent

[77] One could object that LP's requirement is counter-intuitive. To depend on the eventuation of a Discursive Dilemma is like saying I must await the eventuation of a car accident to show I have insurance. Firstly, to respond within the context, you could opt for a counter-factual condition: an agent is autonomous if when facing a Discursive Dilemma, it proceeds to exhibit collective rationality such that group level judgments do not reduce to a majority of individual judgments. Secondly, in section 5.3 I explore a way to conceptualize and articulate these problems in terms of Pareto violations. We might view the capacity for Pareto violations as seen in the evolutionary dynamics preceding the evolution of individuality, as a "real world" correspondence with our theoretical requirements of group agents such that the former lends credence to the latter.

[78] Having said that, we have already seen from other accounts how others, e.g., Rovane and Korsgaard, are prepared to think this way. Drawing on the work of Samir Okasha, I will also take up this hierarchical perspective in regards to rational unity in chapters 4 and 5.

and the degree to which "one" is an agent. In chapter 6, I will suggest that the notion of incentive-compatibility could be exported to similarly describe the biological processes involved in the evolution, development and maintenance of biological individuality.

Before moving on, I will outline and critique LP's own ideas about "developing" agents. I want to further make clear not only questions about agents "agentizing" other agents, but to demonstrate again the presence of the implicit notion many such as LP rely on – that of agents growing and developing into new kinds or degrees or ranks of agency (consistent with the foregoing material in this section). The importance of the particular section of *Group Agency* I focus on below is easy to miss should the reader be overly attuned to the context in which it is presented - that of corporate responsibility. Yet the parenting analogy and vocabulary employed (e.g. "responsibilization")[79] make clear that we are dealing with a deeper conceptual mystery regarding the gray areas between "non-agent" and "agent," or "agent rank 1" and "agent rank 2."

LP make a distinction between "instrumental regulation" and "developmental regulation" that both "bear on the connection between holding agents responsible and action so as to regulate their performance."[80] The distinction itself is unproblematic but the inconsistency with which the concepts are deployed is troubling.

Instrumental regulation pertains to the threat of penalties or promises of rewards in order to shape agents' behavior (e.g. penal regulation of the law). It is in this section that LP express their concern over a false contrast between retributivist versus consequentialist thinking – the former view seen as stressing that one is held responsible and punished according to what they deserve; the latter that one is held responsible and punished in whichever way maximizes the good. To think this way would be a mistake, as it "runs together the issue of how far to hold someone responsible and how far, in light of that responsibility, to impose penalties to regulate" agents' future behavior.[81] To assess whether an agent is responsible is

[79] List and Pettit, *Group Agency*, 157.
[80] List and Pettit, *Group Agency*, 156.
[81] List and Pettit, *Group Agency*, 156. Of course, responsibility for and severity of the act are also importantly distinct. The graveness of the offense alone cannot determine the severity of the punishment. This is because, for example, responsibility may be distributed over a number of people. If the agent's contribution to a grave offense is small, the appropriate (i.e. proportional) punishment should be a function not only of the severity of the offense, but of this low degree of responsibility. Of course, there may also be cases in which the agent is not responsible at all by virtue of the fact that the agent does not meet the three conditions (e.g. an accidental death fails the control condition - a small child who kills fails the judgmental capacity condition).

to *review* the action and determine the agent's fitness to be held responsible (according to LP's trifold criteria: normative significance, judgmental capacity, and control). But to impose a punishment for the purposes of maximizing the good (e.g. making an example of the agent to detour future offenders) is *not a review* of the agent's actions and an evaluation of their fitness to be held responsible. Rather the consequentialist rationale for punishment is prospective in nature and independent from whether one is responsible or not. LP conclude that it would be "grotesque" to hold an agent responsible because it would maximize the good.[82] As I will suggest shortly, LP may be guilty of just such grotesqueness.

In contrast to instrumental regulation, developmental regulation is based on an analogy with parenting. Though children may be yet shy of the age of responsibility, parents may choose to hold children responsible nonetheless for the purpose of "encouraging in the children those very habits that may one day underpin their fitness to be held responsible" that it may "help to induce in them the self-awareness and self-regulation such fitness requires."[83] While LP will acknowledge responsibility and punishment are distinct in the manner explained above, parents are likely to enforce their attitudes with rewards and punishment if they are to be effective.

When LP introduce this concept of "developmental regulation," it is in regard to treating an *agent* as a *morally responsible agent* in an effort to develop the former into the latter. However, later in the same chapter they apply this same "developmental rationale" to treating mere groupings (i.e. collectives) as responsible agents so that they may have reason to incorporate. Examples include religious congregations (distinct from episcopacies), and the descendants of long-dead colonists which may be ascribed "group-level responsibility" (e.g. that they should appropriately feel a kind of collective guilt). Here LP suddenly adopt vague terminology, speaking of "the wider group."[84] Whereas LP were happy in previous sections of the book to talk of whole corporations, states, and churches as group agents, LP now focus on the relationship between the inner and outer elements of these group agents, e.g. the congregation and the episcopacy, a corporation and the board, the citizenry and the government – the former now referred to as merely the group or grouping. This novel nested

[82] List and Pettit, *Group Agency*, 156.
[83] List and Pettit, *Group Agency*, 157.
[84] Consider also the aforementioned "group level responsibility" – a curious phrase with no accompanying explanation or argument to defend what it implies (that for the sake of encouraging a development, a group can justifiably be treated like a morally responsible group agent). What work does the term "level" do in this phrase?

structure seems *ad hoc* - an unexpected revision that is problematic in itself for many reasons. Remaining focused on this idea of the developmental rationale: because the hierarchy more clearly articulated by Jensen (minimal agent – autonomous agent – personal agent) is not explicitly noted or endorsed by LP themselves, it is easy to not at first notice this kind of sleight of hand. Take note of the differences: prodding an agent to develop into a responsible agent versus prodding a collective to become an agent. If I understand LP correctly, developmental regulation only makes sense because there exists an agent which may, by developing habits in response to being held to a higher standard, develop to be capable of meeting the three conditions (normative significance, judgmental capacity, and control). Yet in the case of "groupings," the agent which is said to be developmentally regulated does not yet exist. One might think that the agents that are the target of such prodding are the individuals in the group. However, LP would then face the very problems which their account is designed (in part) to solve (in which case, developmental regulation is either not the solution or it is not developmental regulation). Who is and is not a member of the group? Are all members equally representative of the group? Are all members equally appropriate targets of developmental prodding? Are they a group by virtue of shared goals such that they respond to prodding as responding to incentives?

Curiously, LP seem unbothered by this inconsistency. "It may not strictly be appropriate to hold such a loose grouping responsible, since some of the conditions necessary for fitness to be held responsible are missing."[85] This is not just an understatement – it is misleading. The conditions for being held responsible apply only to group agents, not to groups (or groupings). They continue, "[b]ut holding it responsible may actually prompt the grouping to incorporate and organize against the condemned behavior."[86] The use of "it" in the previous sentence is rather telling.

Here we see that LP are guilty of the grotesqueness described during their discussion of instrumental regulation.

> We can envisage [the developmental rationale's] extension to other cases where the collection that is held responsible falls well short of being a group agent of any kind.

[85] List and Pettit, *Group Agency*, 169.
[86] List and Pettit, *Group Agency*, 169.

> Think of the school group who are told that they will all be held responsible if there is any sign of bullying in their midst; or the loose professional association that is held responsible for the misbehavior of any member; or the neighborhood that is held responsible in the public press when those who live there indulge in certain socially exclusionary acts, say of a racist character; or indeed the generation that is held responsible for the overuse of and potential loss of antibiotics.[87]

Note, while sanctions are implicit in the cases described (e.g. depriving the class of recess, blacklisting the association), there is no explicit talk here of punishment. This is about holding "loose groupings" responsible. In other words, the distinction they caution is preserved. So it is clear that LP hold the very attitude they call grotesque, that is, the attitude of thinking "that whether someone deserves to be held responsible should be determined by whether holding them responsible has desirable consequences."[88]

While on this topic, here are some crucial differences between development in children and "development" of group agents. Unlike children who we want to learn to *become* morally responsible adults because the journey itself is good for them, we would rather have group agents spring into existence fully fit to be held responsible (and perhaps we should be thinking this way when designing regulations). We might prefer group agents that are unfit to be dissolved and replaced by those that are fit rather than opting to mentor or "parent" the unfit as if they were not-yet-fit, as we would with children. Also, unlike children, we do not want group agents, e.g. corporations, to be fit for their own sake because it is good for them, but rather because it is good for us! Human agents must develop from proto-agents – we are not born agents, but we do become agents. A consequence of this means that, for a time, humans are immoral, tyrannical brats (i.e. babies). We would much prefer group agents in society to emerge from the seafoam morally competent.

Let us pull back now from the details. What I hope this shows is that the origin and development of agency is not straightforward. LP's developmental rationale is in line with the Collective Achievement Thesis. Yet it is not clear whether agents are always agents that grow and become "better" in some way, or if there are a series of agent-type categories that

[87] List and Pettit, *Group Agency*, 169.
[88] List and Pettit, *Group Agency*, 156.

social entities progress through. I do not know the answer. But I know that these questions cannot be left in the dark. A way to pursue these questions is suggested in chapter 5. Perhaps a revised account like LPs can offer a suite of more nuanced categories mapping onto the benchmarks of developing agents.

In section 5.1, I will suggest that personhood does not exist on a continuum that all agents can equally develop into. However, the continuum model might have some utility for conceptualizing different types or ranks of non-personal agency achieved more or less to a degree. This might be helpful in meeting what is for many in the group agency literature a primary *desideratum*: holding groups responsible to the degree they are responsible. There is room for plenty of future work amending LP's account to make explicit what they have left vague or implicit (e.g. pluralism about agency and related social categories).

2.4 Collectives

In section 2.2, we took a step back to consider the broader topic of agency (i.e., the "agency" part of "group agency") and different conceptions of agency (e.g. normative). Doing so helps us see further aspects of group agency (e.g. as something achieved, exhibiting potential for disunity, etc.). In this section we will focus on the broader topic of groups (i.e. the "group" part of "group agency").

We categorize groups for our purposes. There are reasons why we might think of any particular group as a corporate body. The concern often taken to be most relevant in the literature is that of responsibility. Considering the growth or development of agents, as we just have in the previous section, implies a range of inchoate collectives merely approximating agency (which, in the spirit of this thesis, might usefully be compared with life forms exhibiting "in-between" individuality, as if midway between an evolutionary transition – see section 3.3.1). Stephanie Collins provides a useful taxonomy for different types of collectives.[89]

Like others such as LP, French, and Dennett, Collins begins her treatment of collectives by pointing to the way we use language. She concludes that similarly worded responsibility

[89] Collins, "Collective responsibility gaps."

ascriptions might be directed at any one of three different types of collectives: diffuse, teleological, and agential collectives (the latter being a group agent). According to Collins, these crosscut with three types of responsibility (visually represented in the Table 2.1: causal, moral, and prospective).

		Responsibility		
		Causal responsibility	**Moral responsibility**	**Prospective responsibility**
Collective	Diffuse collective	e.g. '19th century industrialists affected global warming.'	e.g. 'Present-day motor companies are culpable for causing global warming.'	e.g. 'Present-day motor companies have a duty to ameliorate global warming.'
	Teleological collective	e.g. 'The fossil fuels lobby causes lax government policy.'	e.g. 'The fossil fuels lobby is to blame for global warming.'	e.g. 'The fossil fuels lobby would do the right thing by ceasing to exist.'
	Agential collective	e.g. 'BP had a role in the Deep Horizon oil spill.'	e.g. 'BP is to be praised for its efforts.'	e.g. 'BP has an obligation to clean up its oil spill.'

Table 2.1: Mapping types of responsibility to types of collectives[90]

Diffuse collectives may be collections that LP would call "mere collections" along with groups that share a property in common, e.g. all blued eyed people.[91] However, Collins has in mind collections more likely to be the subject of your responsibility ascriptions, such as "affluent consumers" or "tech companies." Teleological collectives are those in which individuals work toward shared goals but have no decision procedure. An example of the latter is a spontaneously congregated group of beach-goers cooperating to assist a stranded whale.

Causal responsibility can be attributed in events devoid of action, such as a tornado being responsible for the destruction of one's home. Moral responsibility involves causal

[90] Reproduced from Collins, "Collective Responsibility Gaps," 945.
[91] List and Pettit, *Group Agency*, 31.

responsibility but also backward-looking praise or blame. Prospective responsibility refers to forward-looking obligations. Prospective responsibility means having "a moral reason that is presumptively decisive in decision-making."[92] Causal responsibility and moral responsibility both track responsibility to past events. Moral responsibility and prospective responsibility both involve moral judgments and therefore must involve actions.

Diffuse collectives can be causally responsible. Even if we do not know the specific individuals in the diffuse collective, "the grouping has the right level of detail" to provide the relevant explanation. Following List and Menzies, Collins says that the causal power of a diffuse collective is "realization insensitive."[93] While we can attribute causal responsibility to diffuse collectives, we cannot attribute moral or prospective responsibility.

For there to be moral responsibility for an outcome, there must *have been a past volition that did* (in some way) cause the outcome. For there to be prospective responsibility for an outcome, there must be *the possibility of a future volition, produced via moral reasoning, which could* (in some way) cause the outcome.

Collins does not mention that what is necessary here is not sufficient. Just because a child or psychopath can be shown to have wanted to kick you in the head, and that the desire to kick you is what caused it, does not make them morally responsible. This is because they are not fit for moral responsibility. I think LP's articulation of "fitness to be held responsible" would be a useful supplement if not a better alternative (and thus one might have good reason to revise Collins' taxonomy appropriately).

Teleological collectives are trickier. Because these collectives share goals and act responsively to one another in respect to their collective efforts, Collins suggests the best option is to say that they lack obligations and therefore cannot be blamed at the group level. Individuals within the collective may be blamed or praised (e.g., for their contributions to the collective action, for their membership in the "harm-producing group"), but the collective cannot be blamed or praised. Of course, teleological collectives cannot also bear prospective responsibility, which is reserved for those with the capacity for group-level moral decision-

[92] Collins, "Collective Responsibility Gaps," 945.
[93] Christian List and Peter Menzies, "Nonreductive physicalism and the limits of the exclusion principle," *The Journal of Philosophy* 106, no. 9 (2009): 496.

making.

Agential collectives can be held prospectively responsible (and therefore also morally and causally) as they are moral agents.

Collins' taxonomy offers an alternative to talk of collections versus groups, or groups versus group agents, or group agents versus group persons (or "fully fledged" group agents). Collins maps different types of collectives to different types of responsibility in order to make sense of responsibility attributions in our everyday language. (In particular she is concerned with responsibility gaps and addresses potential gaps in a more detailed way than LP do when citing their motivation for the account of group agency they put forward). Interestingly, I think Collins' taxonomy gives us an alternative to talking about whether our ascriptions are shorthand versus actual as when we consider taking up the intentional stance. What may be otherwise considered mere shorthand or metonymy may alternatively be considered a type of collective (diffuse) with a label (e.g. "right wingers") that actually refers, even if we do not know which individuals are precisely implicated. What value is there in two ways of saying the same thing? It is at least minimally valuable in that talk of "mere shorthand" or "metaphor" is easier to dismiss altogether, when we might otherwise be able to say something about them. For example, if a statement implicating a "mere collection" is thought of as shorthand, we will not think to further consider whether the collection would be responsible. However, if we think of the statement referring to a kind of collective, a diffuse collective, we can say that perhaps the collective is causally responsible, or negatively, rule out or find false, statements that would attribute moral or prospective responsibility. In short, Collins' taxonomy of types of collectives (which could have been variously named e.g., "group agents," "groups," and "groupings") might prove to be useful in providing a bit of clarity when talking about collectives with appreciable differences. I will put her terms to work in section 5.1.5.

We should note there are different ways to group useful categories. For example, while Collins speaks in terms of different types of both collectives and responsibility, Pettit talks about three "levels at which the agency of individuals may combine:"[94]

[94] Philip Pettit, "Corporate Agency: The Lesson of the Discursive Dilemma," in *The Routledge Handbook of Collective Intentionality*, eds. Marija Jankovic and Kirk Ludwig (New York: Routledge, 2018): 249.

1. Common Effect - an effect of various individual actions from agents pursuing their own goals (buyers and sellers affecting prices)
2. Single action - Individuals act with shared or joint agency to together achieve a goal or bring about a desired change (e.g. good Samaritans coordinating their actions at the site of a car accident to save an injured man's life)
3. Group Agency - When individual agents constitute a single agent.

Categorizing variously types of actions, collectives, causes, or responsibility may have slightly different connotations and implications. For example, invoking the concept of "levels" may require spelling out exactly what that means, with potential for nuanced disagreement in various positions (e.g. disagreeing that what happens at level 2, however similar to level 1, cannot count as an *action*, or *cause* or whatever).[95] Therefore, the extent to which we can be precise will often be limited by the scope of our particular projects (e.g. focused on explaining action rather than offering criteria for agency). A larger project than this could tackle the wide web matching contexts with concepts, accounting for the prior commitments of each, perhaps drawing on a wider range of ideas from different sub-disciplines (e.g. emergence, supervenience, "top-down" causality, networks, etc.). I take this to be further evidence that this topic is ripe for many "bridge-builders" to contribute to the literature.

In chapter 5, I will suggest a potential application for taxonomies like those of Collins, and a model by which we might create and organize similar taxonomies. LP have arguably provided a well-argued account of group agents as distinct from short-lived, unorganized groups. When groups exhibit agency they can be held to account. This is a prime motivation for LP and others in the literature. But this does not change the fact that many collective actions of unorganized groups can have enormous effects on the world. In other words, some of the largest responsibility deficits might be found in borderline cases (e.g., Collins'

[95] On my view, articulating the varying scenarios of collective action in terms of "levels" as Pettit does is confusing. Throughout this thesis my usage of "levels" corresponds to nested hierarchies of individuals and groups. I think Pettit meant to express the degrees of coordination and unity in which collections of individuals exhibit in their actions. Alternatively, we can say he has distinguished individual action, joint action, and group agency. This point demonstrates the different angles from which one can approach the wider topic: focusing on the combined actions, the effect of the combined actions, or the features of the collective. These varying approaches are represented by the related literatures of e.g., shared agency, joint intentionality, and collective responsibility.

teleological collectives). I do not think this problem will go away nor will we find an easy answer, but in section 5.1.5 I will offer a suggestion.

Chapter 3
Biological Individuality

This chapter explores the topic of biological individuality currently debated in the philosophy of biology literature. In Section 1, I give an overview of the topic. In Section 2, I provide a more detailed snapshot of the current state of the literature, highlighting some commonly identified problems and recent trends. In Section 3, I discuss the evolution of individuality. In Section 4, I outline the accounts of biological individuality offered by Ellen Clarke, Thomas Pradeu, and Maureen O'Malley. In Section 5, I consider some cases that put pressure on our thinking about biological individuality. In Section 6, I consider a pluralistic approach to biological individuality, featuring the accounts of Peter Godfrey-Smith, Jack Wilson, and James DiFrisco. In Section 7, I offer a brief summary of the chapter and a glimpse at how material from this chapter will inform subsequent chapters.

3.1 Biological Individuality

The living world is full of discrete entities we can point out, individuate, and speak of with reasonable precision. We can easily track the movements of lobsters and follow them around. We can easily make the distinction between a grasshopper and its environment. We know that when a whale gives birth, the new whale is a new individual and not simply more of the first whale (nor whale excrement).

Other cases are not so clear cut. Some *appear* clear cut, but upon closer inspection, are vexing even to – perhaps especially to – experienced biologists. What phenomenally appear as separate and distinct biological entities may be part of, or heavily depend on, other biological entities. Consider a few examples.

A virus is a protein-coated genetic segment that can replicate and spread, but only by utilizing the reproductive machinery of a host cell. Quaking aspen trees appear as separate, discrete units above ground but are connected beneath the ground by a vast root system.[1] The tongue-eating louse is a parasitic isopod that severs and removes a fish's tongue, then itself

[1] Jeffry B. Mitton and Michael C. Grant, "Genetic variation and the natural history of quaking aspen," *Bioscience* 46, no. 1 (1996): 25-31.

becomes the fish's new tongue.[2] Similarly, a riftia tubeworm's digestive tract is not merely colonized but *replaced* by symbiotic bacteria such that for nearly all but the earliest stages of the tubeworm's life, its gut *is* environmentally-acquired bacteria.[3] A slime mold is a population of genetically heterogeneous single celled amoebae that behave(s) like a loosely organized colony at times, and a unified organism at other times (in which some amoeba sacrifice themselves in order to facilitate the distribution of spores).[4] Depending on the season, any given aphid might be the result of asexual or sexual reproduction – with parent and offspring genetically clonal or genetically distinct, respectively.[5] Some plants undergo a metagenic life-cycle, successively reproducing what are often very different looking specimens (i.e. phenomenally different kinds of individual): gametophytes (haploid) and sporophytes (diploid).[6] Some social insect colonies are functionally integrated such that they operate much like animals – with castes likened to organs – such that they are subject to natural selection at the colony level, in evolutionary competition with other colonies.[7]

The biological world is full of edge cases in which we are uncertain whether we are observing individuals or groups, reproduction or growth. The answers will vary depending on what we want to know. That is, we will get different answers to questions that implicate different criteria for individuality. Daniel Janzen famously described patches of genetically identical dandelions as a trunkless, rootless tree.[8] Obviously, Janzen's criteria for individuality lacked a requirement for physical contiguity or for a boundary. Rather, he was only considering the genetically unique product of sexual reproduction however scattered that product may be. Hence, Janzen strategically invokes the concept of a tree so that we hold in our mind, with conceptual force equally binding as branches, the historical and genetic

[2] Richard C. Brusca and Matthew R. Gilligan, "Tongue replacement in a marine fish (Lutjanus guttatus) by a parasitic isopod (Crustacea: Isopoda)," *Copeia* 1983, no. 3 (1983): 813-816.
[3] "Symbiosis" refers to the long-term interaction of different species. Symbiotic arrangements can be mutualistic, parasitic, or commensal.
[4] John Tyler Bonner, "Evidence for the sorting out of cells in the development of the cellular slime molds," *Proceedings of the National Academy of Sciences of the United States of America* 45, no. 3 (1959): 379-84. Leo W. Buss, *The Evolution of Individuality* (Princeton: Princeton University Press, 1987).
[5] Daniel H. Janzen, "What are dandelions and aphids?" *The American Naturalist* 111, no. 979 (1977): 586-589.
[6] "Metagenic" refers to reproducing alternately, asexually (sporophytes) and sexually (gametophytes). This is sometimes referred to as the alternation of generations.
[7] David Sloan Wilson, "Altruism and organism: Disentangling the themes of multilevel selection theory," *The American Naturalist* 150, no. S1 (1997): s122-S134; Edward O. Wilson, *Sociobiology: The New Synthesis* (Cambridge: Harvard University Press, 1975); Edward O. Wilson and Bert Hölldobler, *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies* (New York: W.W. Norton, 2009).
[8] Janzen, "What Are Dandelions and Aphids?"

connections between non-contiguous parts, i.e. flowers. Janzen's sense of individuality is useful for anyone who wants to track or quantify the products of sexual reproduction.

By contrast, Janzen's sense is less useful for other purposes. A harvester of dandelions could pull 30 stems from the ground and look to the bounty of their basket unbothered by their ignorance as to how many of Janzen's "individuals" they have gathered (1? 5? 30?). Similarly, for everyday purposes in our world, two physiologically discrete human beings who happen to be the product of a single zygote (i.e. identical twins) would never be thought to comprise a single individual.[9] Likewise, a lumberjack tasked with cutting down 50 aspens does not need to examine the root system to do so. But the problem is not as easy as reconciling the intuition of laypersons with the technical language of specialists (e.g., between lumberjacks and arborists). Specialists themselves employ many concepts of biological individuality – many which should not be used interchangeably. A diverse range of biological entities have been called individuals: organisms,[10] organelles,[11] genes,[12] species[13] trait groups,[14] colonies,[15] cells,[16] holobionts,[17] communities,[18] clades, and all of life on Earth,[19] among others. Some will see this list as motivating the case for pluralism about biological individuality.[20]

---

[9] Even when genetic identity is the basis for a project, e.g. comparing the traits of a returning astronaut with his terrestrially bound twin, the qualities they share are seen as *aspects* of two individuals. More interesting cases are those like e.g. the Hensel twins, conjoined twins which can simply, if crudely, be described as what appears physiologically to be a two headed human. The latter is a case of biological contiguity without individuality (in the relevant sense).

[10] David C. Queller and Joan E. Strassmann, "Beyond society: the evolution of organismality," *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, no. 1533 (2009): 3143-3155.

[11] Thomas Pradeu, "Organisms or biological individuals? Combining physiological and evolutionary individuality," *Biology & Philosophy* 31, no. 6 (2016): 797-817.

[12] David L. Hull, "A matter of individuality," *Philosophy of science* 45, no. 3 (1978): 335-360.

[13] Michael T. Ghiselin, "A radical solution to the species problem," *Systematic Biology* 23, no. 4 (1974): 536-544.

[14] David Sloan Wilson, "Structured demes and the evolution of group-advantageous traits," *The American Naturalist* 111, no. 977 (1977): 157-185.

[15] Matthew Haber, "Colonies are individuals: revisiting the superorganism revival," In *From groups to individuals: evolution and emerging individuality*, eds. Frédéric Bouchard and Philippe Huneman (Cambridge: MIT Press, 2013), 195-217. See also, William M. Wheeler, "The ant-colony as an organism," *Journal of Morphology* 22, no. 2 (1911): 307-325, in which individuality is an organismal quality.

[16] Richard Dawkins, *The Extended Phenotype: The Long Reach of the Gene* (New York: Oxford, 1982), 260.

[17] Javier Suárez and Adrian Stencel, "A part-dependent account of biological individuality: Why holobionts are individuals and ecosystems simultaneously," *Biological Reviews* 95, no. 5 (2020): 1308-1324.

[18] Marc Ereshefsky and Makmiller Pedroso, "Rethinking evolutionary individuality," *Proceedings of the National Academy of Sciences* 112, no. 33 (2015): 10126-10132.

[19] Carlos Mariscal and W. Ford Doolittle, "Life and life only: a radical alternative to life definitionism," *Synthese* 197, no. 7 (2020): 2975-2989.

[20] Cf. Pepper and Herron's organism-centered approach; they suggest the range of usages might be thought more simply to apply to organisms, parts of organisms, or groups of organisms. John W. Pepper and Matthew D. Herron, "Does biology need an organism concept?" *Biological Reviews* 83, no. 4 (2008): 621-627.

So, there are different ways to carve up the biological world and a range of different entities we call "individual." Is this a problem? Many will argue emphatically that it is undoubtedly a problem.[21] Our best theories often take the existence of well delineated individuals for granted (e.g. the theory of natural selection). Without being able to identify individuals, how might one carry out the following?

- conduct an ecological census
- distinguish one creature from another
- distinguish reproduction from growth
- calculate the fitness of an organism (i.e. count offspring)
- tell the difference between a creature and a mere part of a creature
- differentiate a new life form from a successive stage in the same life cycle

Questions like these show what turns on the problem of biological individuality.[22]

But before exploring the debate about what it means to be a biological individual, we must first be clear on the more general point of what it means to be an individual biologically. In metaphysics, individuality is normally understood as an ontological category. As an ontological category there are features of individuality which cannot admit of degrees; James DiFrisco cites examples such as being necessarily uniquely located in space-time, being a particular, concreteness, persistence, and non-dependence.[23]

However, there is a vast landscape of not-quite-individuality in biology.[24] This is because biological individuality is a *scientific* kind or class.[25] One can be a kind of biological

[21] Ellen Clarke, "The problem of biological individuality," *Biological theory* 5, no. 4 (2010): 312-325.
[22] Some (e.g. Karen Kovaka) have expressed doubt that the problem of biological individuality is a *problem for biologists*. However, the justification for drawing lessons from the individuality debate for the group agency literature in this thesis rests on the former debate's status as a *philosophical* debate, not on its ultimate relevance for practicing biologists. See section 5.2.6. Karen Kovaka, "Biological individuality and scientific practice," *Philosophy of Science* 82, no. 5 (2015): 1092-1103.
[23] James DiFrisco, "Kinds of Biological Individuals: Sortals, Projectibility, and Selection," The *British Journal for the philosophy of science 70, no. 3* (Sep 2019): 850.
[24] Stéphane Chauvier, "Individuality and aggregativity," *Philosophy, Theory, and Practice in Biology* 9, no. 11 (2017): 1-14.
[25] I use "kind" synonymously with "class." In this thesis I will not say anything about natural kinds and the various stances on whether and which individuals count as natural kinds.

individual to a degree but still remain fully an individual in the ontological sense.[26] Many philosophers of biology hold that biological individuality is best considered this way - as something which admits of degrees.[27] (There are those who prefer to talk of "organismality" on the grounds that "individuality" should remain a categorical concept that does not admit of degrees.[28] However, preserving semantic consistency with usage in metaphysics comes at the cost of adding further ambiguity within the biological literature itself, i.e. regarding the term "organism").

Something can be more or less an individual based on possessing more or less of what makes an individual (whatever we take that to be, e.g. levels of cooperation, efficacious policing mechanisms). So, while many share this idea of degrees of individuality, they may disagree on what exactly makes an individual more or less an individual. We will now turn to these disagreements (briefly in section 3.2, with some detail highlighted later in section 3.4).

3.2 The Problem of Biological Individuality – Current State of the Literature

What does it mean to be a biological individual? Lidgard and Nyhart describe the puzzle of biological individuality in terms of a "problem space," one that changes and evolves as the subject is studied.[29] One sets out to individuate some set of biological entities, and soon after, can reflect on the resulting individuality concept as context dependent (suited to the purposes of the project and environment, which will differ from those of other researchers). Clarke and Okasha distinguish the problems of vagueness and ambiguity.[30] The problem of vagueness refers to confusing borderline cases in which we struggle to delimit particular individuals. The problem of ambiguity refers to the various conceptions of individuality as informed by the range of definitional criteria on offer. The two are related for those who think that the

[26] DiFrisco, "Kinds of Biological Individuals."
[27] E.g. Richard Michod, *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality* (Princeton: Princeton University Press, 1999); Bernabé Santelices, "How many kinds of individual are there?" *Trends in ecology & evolution* 14, no. 4 (1999): 152-155; Matthew Herron et al., "Cellular differentiation and individuality in the 'minor' multicellular taxa," *Biological Reviews* 88, no. 4 (2013): 844-861; Peter Godfrey-Smith, Peter, *Darwinian populations and natural selection* (Oxford University Press, 2009); Ellen Clarke, "The multiple realizability of biological individuals," *The Journal of Philosophy* 110, no. 8 (2013): 413-435.
[28] E.g. Subrena Smith, "Organisms as persisters," *Philosophy, Theory, and Practice in Biology* 9, no. 14 (2017): 2.
[29] Scott Lidgard and Lynn K. Nyhart, eds., *Biological individuality: Integrating scientific, philosophical, and historical perspectives* (University of Chicago Press, 2017), 25.
[30] Ellen Clarke and Samir Okasha, "Species and organisms: What are the problems," in Bouchard and Huneman, *From groups to individuals*, 55-75.

more distinct criteria are fulfilled, the higher the degree of individuality (of a particular biological entity).

Various accounts offer different criteria. To the extent that different projects have different priorities, then individuality concepts guided by what is important in different contexts will differ accordingly.[31] Several philosophers have already reviewed the accounts on offer to convincingly show that there can be no consistent account of biological individuality that meets all of the candidate criteria.[32] The problem of biological individuality is about an unsatisfying absence of theoretical unity with no consensus on how to proceed.

The features variously considered to be required for, characteristic of, or paramount to individuality include: functional integration,[33] reproduction,[34] genetic homogeneity,[35] cooperation among intra-individual constituents,[36] reproductive division of labor,[37] capacity to undergo natural selection,[38] conflict-suppressing mechanisms,[39] physical contiguity,[40] adaptations,[41] life cycle starting with a single-cell (or otherwise "bottlenecked"),[42] immune response,[43] and others.[44]

Several philosophers have done the neat work of demonstrating how various accounts of individuality are unable to account for the range of criteria thought to be necessary for individuality.[45] In an influential paper, Ellen Clarke concludes that verdicts on whether any

[31] To complicate matters, agreeing on what conditions must be met to count as an individual does not mean agreeing on what counts as meeting those conditions, or whether they have in fact been met, or to what extent, or where thresholds are placed.
[32] E.g. Clarke, "The problem of biological individuality," DiFrisco, "Kinds of biological individuals."
[33] E.g. David Sloan Wilson and Elliott Sober, "Reintroducing group selection to the human behavioral sciences," *Behavioral and brain sciences* 17, no. 4 (1994): 585-608.
[34] E.g. James Griesemer, "Development, culture, and the units of inheritance," *Philosophy of science* 67, no. 3 (2000): S348-S368.
[35] E.g. John L. Harper, *Population Biology of Plants* (New York: Academic Press, 1977).
[36] E.g. Queller and Strassman, "Beyond Society."
[37] Leo W. Buss, "Evolution, development, and the units of selection," *Proceedings of the National Academy of Sciences* 80, no. 5 (1983): 1387-1391.
[38] E.g. Scott F. Gilbert et al., "A symbiotic view of life: we have never been individuals," *The Quarterly review of biology* 87, no. 4 (2012): 325-341.
[39] E.g. Michod, *Darwinian Dynamics.*
[40] E.g. Julian Huxley, *The Individual in the Animal Kingdom* (New York: University Press, 1912).
[41] E.g. Henri J. Folse III and Joan Roughgarden, "What is an individual organism? A multilevel selection perspective," *The Quarterly review of biology* 85, no. 4 (2010): 447-472.
[42] E.g. Dawkins, Richard. *The extended phenotype*.
[43] E.g. Thomas Pradeu, *The limits of the self: immunology and biological identity* (Oxford University Press, 2011).
[44] For an impressive historical survey see Lidgard & Nyhart, *Biological individuality.*
[45] E.g., Clarke, "The problem of biological individuality;" DiFrisco, "Kinds of Biological Individuals."

given biological entity is an individual will differ.[46] For example, if sexual reproduction is required for individuality, then neither a bee colony nor a bacterium are individuals. But if the sequestration of the germline is what is required for biological individuality, a bee colony is an individual while the bacterium is not. If spatial boundedness is required, then the bacterium is an individual while the bee colony is not. If a bottle-neck life cycle is what is required, both are individuals.

Implicitly or explicitly, the chosen criteria are often a claim about what is *important*. For example, evolutionary accounts of biological individuality stress the centrality of populations undergoing natural selection. To be an individual on an evolutionary account is to be a member of a Darwinian population.[47] Consistent with important work that highlights the evolution of individuality itself,[48] some philosophers point to the evolved mechanisms that "maintain" individuality[49] - mechanisms that may vary in their material and structural realizations.[50]

In contrast, organizational accounts emphasize the functional integration and persistence of biological individuals.[51] Though functional organization might be the product of evolution, it need not be in principle. On this understanding, biological individuals need not form a lineage (though their parts may).[52] Some emphasize the integrity of individuals achieved by high levels of cooperation and low levels of conflict within.[53] Others emphasize the presence of constraints ensuring "organizational closure."[54]

These two broad categories have given rise to a recent literature that posits two main types of biological individual: evolutionary and physiological.[55] Evolutionary individuals reproduce

---

[46] Clarke, "The problem of biological individuality."
[47] Godfrey-Smith, *Darwinian populations and natural selection*.
[48] Leo W. Buss, *The Evolution of Individuality*; John Maynard Smith and Eors Szathmary, *The Major Transitions in Evolution* (New York: Freeman 1995).
[49] Michod, *Darwinian Dynamics.*
[50] Clarke, "The multiple realizability of biological individuals."
[51] David Sloan Wilson, "Biological communities as functionally organized units," *Ecology* 78, no. 7 (1997): 2018-2024.
[52] Haber, "Colonies are individuals."
[53] Queller & Strassmann, "Beyond society."
[54] E.g. James DiFrisco and Matteo Mossio, "Diachronic identity in complex life cycles: an organizational perspective," in *Biological Identity: Perspectives from Metaphysics and the Philosophy of Biology*, eds., A.S. Meincke and John Dupré (New York, Routledge: 2021), 177-199.
[55] Pradeu, "Organisms or biological individuals?" See also Robert A. Wilson and Matthew J. Barker, "Biological Individuals," *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), ed., Edward N. Zalta, https://plato.stanford.edu/archives/fall2019/entries/biology-individual.

and form lineages. Physiological individuals persist as metabolic collaborations of their parts.[56] The reproductive and metabolic capacities that characterize evolutionary and physiological individuals, respectively, coincide in what we often consider to be paradigmatic biological individuals. Some find overlapping cases an intuitive place to apply the folk term, "organism"[57] (while others would prefer that this term disappear[58]).

Note that on this understanding not all biological individuals are living. A gene or chromosome might be an individual in the evolutionary sense because it meets natural selection criteria (e.g. heritable variation in fitness).[59] Our intuitions about what is living often corresponds with a functioning metabolism, not just, or even, the capacity to reproduce. As we will see, cases like those of genes and viruses are what motivate the use of more nuanced categories by pluralists (e.g. "scaffolded reproducers")[60] to describe those evolutionary individuals which cannot reproduce on their own; such entities require external metabolic machinery (e.g. genes, viruses). Similarly, we do not intuitively take some symbiotic partnerships and colonies to be alive as wholes, but rather to be only comprised of multiple living things (although our intuitions differ from case to case). Trickier still, in the cases of the earth that makes up termite mounds and the calcium deposits that make up coral reefs, we must consider whether non-living abiotic aspects of the environment are properly considered parts of a larger biological individual.[61]

To recap, we have introduced the problem of biological individuality. We have noted a range of candidate criteria or definitional features. Several philosophers have documented the challenge within the literature of offering an account that can accommodate the entirety of this range; the criteria draw different boundaries around the living world (i.e., do or do not count certain entities as individuals). The current trend within the literature is to think of

[56] Haber "Colonies are individuals."
[57] E.g. Pradeu, *Limits of the self*; Godfrey-Smith, *Darwinian Populations.*
[58] E.g., Haber "Colonies are individuals."
[59] Richard C. Lewontin, "The units of selection," *Annual review of ecology and systematics* 1, no. 1 (1970): 1-18.
[60] Originally Kim Sterelny's term, but used by Godfrey-Smith in the account featured in chapter 5.
[61] It is worth flagging that the relations between talk of biological individuals and what it means to be alive or living is, in my opinion, an under explored intersection of interests in the literature; addressing both problems simultaneously would make for interesting work (with potential for, in the spirit of this thesis, each subject informing the other). Cf. Robert Wilson who focuses on "agents," not individuals, gives explicit attention to the set of "living agents." Robert A. Wilson, *Genes and the agents of life: the individual in the fragile sciences biology* (Cambridge: Cambridge University Press, 2005); Lynn Margulis and Dorion Sagan, "What is Life?" (New York: Simon and Schuster: 1995).

"biological individuals" as an overarching category in which other individuality concepts nest. Broadly speaking, these nuanced concepts are often guided by considerations of evolution and reproduction on the one hand, and considerations of organization and metabolic persistence on the other hand, with paradigm cases of individuality exhibiting readily identifiable features of both.

3.3. The Evolution of Individuality

In the next section (3.4), we will look to some specific accounts. Doing so will help not only to get a closer look at the debate, but also draw out the background concepts in play. However, some of that background information should be presented up front. It will be useful to address the evolution of individuality itself, which will help contextualize many of the concerns featured in a range of accounts (not only evolutionary accounts), as well as provide the conceptual tools necessary to appreciate some of the "fuzzy" cases of biological individuality we will encounter (e.g., is any given borderline case a product of entities that exhibit few degrees of individuality in their current evolutionary state,[62] or the result of a non-evolutionary metabolic collaboration among evolutionary individuals?). Lastly, I will highlight (but not fully develop here) two themes which are potentially exportable for wider consideration within the context of agency and group agency. This discussion lays some groundwork for chapters 4 and 5.

3.3.1 Evolutionary Transitions in Individuality and Multi-Selection Theory

In the moral and political domain, when we want to understand the origin of groups, we look to individuals. Individuals have goals they want to achieve. They then form a group for the purpose of bringing about those ends.

In the domain of evolutionary biology, however, we look to groups to explain the origin of individuality.[63] Individuality is an evolved character.[64] This sounds at once completely obvious and extremely counter-intuitive. On the one hand, it seems counter-intuitive due to

[62] See e.g., Chauvier, "Individuality and aggregativity."
[63] Buss, *The Evolution of Individuality*; Richard Michod, "Evolution of individuality during the transition from unicellular to multicellular life," *Proceedings of the National Academy of Sciences* 104, no. S1 (2007): 8613-8618; Michod, *Darwinian Dynamics;* Maynard-Smith and Szathmary, "Major Transitions in Evolution."
[64] Buss, *The Evolution of Individuality*.

how we might normally conceptualize mereological composition as a sort of building up – we think of individuals as prior to groups. On the other hand, it seems obvious in light of the fact that all the living things we observe today have evolved from the aggregation of less complex forms, going all the way back to the first self-replicating molecules.

Two ancillary notions seem to make the counter-intuitive obvious: time and levels ("levels" in the sense of a nested hierarchy). An individual at one level may have other types of individuals as its parts. This hierarchy is the product of evolution. Therefore, our intuition that individuals are somehow "prior" to groups is consistent with the fact that individuality is a derived character. A biological individual is in some sense a community of other individuals. That's true. It is also true that the individuality of those constitutive component individuals is something achieved by the historical and evolutionary predecessors of the larger individuals they compose.

Biological life often presents in clearly organized compositional hierarchies. Natural selection may operate at these different hierarchical levels. In fact, selection is active even at levels in which entities which make up populations may not themselves be living (though they will be "Darwinian,"[65] e.g. viruses, genes). But to recognize that natural selection operates at different levels is not itself satisfying when thinking about transitions in individuality – we want to know why there are levels in the first place! Our curiosity does not (and should not) end with an answer to a question about which level selection is currently acting on. We want to know: which mechanisms gave rise to the biological hierarchy that is itself a product of evolution?[66] Perhaps not unlike geology, there is a story in the layers; and the story is not in the existence of multiple levels, but in the *levelling up*.

We can pinpoint key moments in the expansion of the biological hierarchy, referred to as "evolutionary transitions in individuality."[67] A transition in individuality is when previously free-living entities combine into an organized whole within which they henceforth function as parts, only reproducing as part of the larger whole (i.e. what was formerly reproduction, is

[65] Following Peter Godfrey-Smith.
[66] Samir Okasha, *Evolution and the Levels of Selection* (Oxford University Press, 2006), 16.
[67] Michod, *Darwinian Dynamics.* Many will prefer the more general term, "major transitions" following Maynard Smith and Szathmary, even though the latter picks out a wider set of evolutionary events, not all of which resulted in the addition of a new level to the compositional hierarchy of living things (i.e. but resulted in a new way in which populations evolve and information is transmitted).

now growth).[68] Richard Michod articulates the latter aspect of this transition in terms of fitness being transferred or exported to a higher level – the new aggregate entity.[69] These novel entities exhibiting heritable fitness variance are then subject to selection as wholes, competing at their own level of interaction. The current living world is a testament to the history of such transitions. For example, at some point in history, self-replicating molecules that organized themselves into clusters began reproducing as clusters. Those clusters organized themselves into clusters of clusters, and again, were selected to reproduce as wholes (i.e. clusters-of-clusters) rather than independently. That is, for whatever reason, clustering (and clustering of clusters and so on) conferred evolutionary benefits superior to that of not clustering. Clusterers were selected because of their clustering properties and as a result, clusterers produced a more fruitful lineage. Surviving and reproducing as a whole means that, however the details played out, the whole became solidified, and itself subject to natural selection as it competed with other emerging solidified wholes. This progression of what I've been loosely referring to in the abstract as "clustering" is meant to apply to evolutionary transitions such as that from replicating molecules to complexes of replicators; from those complexes to prokaryotic cells; from prokaryotic cells to eukaryotic cells; from eukaryotic cells to multicellular organisms; and from multicellular organisms to groups of multicellular organisms. (Though the inclusion of the latter has been a source of debate over the last century).

How do self-replicating entities organize into clusters? There is a deep mystery as to how new individuals arise from a collection of lower-level individuals. Sometimes we imagine a change within the population from competitive to cooperative. Entities which in previous states of evolution existed independently become tightly integrated, having sufficiently quelled conflict that would have otherwise made symbiosis impossible. As natural selection acts on the integrated whole, the cooperative venture increasingly takes the shape of a paradigm individual. Most find plausible the idea that a sub-population of altruistic organisms might outcompete a sub-population of selfish organisms and that as a result, group-level selection may occur, with selection henceforth refining and solidifying such dominating groups.[70] Alternatively or in addition (i.e. for different types of transitions), we

[68] Maynard-Smith and Szathmary say something similar but they use "replicate" which would only apply to some transitions. Maynard-Smith and Szathmary, "Major Transitions in Evolution," 8.
[69] Michod, *Darwinian Dynamics*.
[70] See Travisano et al., "Origins of multicellular evolvability in snowflake yeast," *Nature communications* 6, no. 1 (2015): 1-9.

can think of transitions being the result of accidents, such as one cell "swallowing" another, then somehow reproducing as a whole thereafter, outcompeting others due to its size. A similar story might be told about "sticky" cells, clinging together post-mitosis, henceforth multiplying in the aggregate.

What we currently consider parts of individuals were once themselves individuals. Mitochondria are now considered organelles, but at one time were free-living single-celled organisms. That is, mitochondria have been through a sort of journey, from free living, to endosymbionts, to organelles. Such events in evolution's history are rare, but not necessarily singular. Multicellularity has evolved separately on around two dozen occasions in different lineages.[71] This means there are a number of historical instances in which new, more complex organisms came about via the coming together of simpler organisms.[72]

The above hypothetical examples of "swallowing" and "stickiness"[73] feature organisms and cells, but we could further imagine other similar cases with replicating molecules and complexes of such molecules. The mystery is a general one, with more specific sub-mysteries featured in each instance, at each level, with each type of successive transition.

Telling these kinds of stories allows us to surmise how these events may have happened in life's history. But once one begins to more carefully analyze cases of multi-level selection, one is dumbfounded by fundamental problems with the conditions and criteria.[74] Once we consider what it might take for evolutionary transitions in individuality to actually happen we

---

[71] Ruiz-Mirazo et al., "Organisms and their place in biology," *Theory in biosciences* 119, no. 3 (2000): 209-233.
[72] Maynard Smith & Szathmáry *The Major Transitions in Evolution;* Michod, *Darwinian Dynamics;* David C. Queller, "Relatedness and the fraternal major transitions," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 355, no. 1403 (2000): 1647-1655; Antonelli et. al., "Theories and models in symbiogenesis," *Nonlinear Analysis: Real World Applications* 4, no. 5 (2003): 743-753; Richard Michod and Matthew Herron, "Cooperation and conflict during evolutionary transitions in individuality," *Journal of evolutionary biology* 19, no. 5 (2006): 1406-1409; Susse Kirkelund *et al.*, "Evolution of species interactions in a biofilm community," *Nature* 445, 533–536 (2007). While we can point to modern biological individuals on Earth (e.g. rabbits) as having contingently evolved following these special historical transition events in which biological entities came to *evolve in new ways* (not just evolve new traits), it does not rule out the possibility that "playing back the tape" would also expand the biological hierarchy of individuality. This was Stephen Jay Gould's way pumping the intuition about evolutionary counterfactuals that we would nowadays articulate in terms of running complex simulations that would produce very different outcomes sensitive to initial conditions.
[73] Stickiness is an example of a "fraternal transition," and the swallowing story can be either, but e.g. in the case of eukaryotic cells in mitochondria, is an example of an "egalitarian transition."
[74] Multi-level selection theorists include John Damuth and I. Lorraine Heisler, "Alternative formulations of multilevel selection," *Biology and Philosophy* 3, no. 4 (1988): 407-430; Elliott Sober and David Sloan Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Cambridge: Harvard University Press, 1998); Okasha, *Evolution and the levels of selection.*

realize many of our ready-to-hand evolutionary explanations beg the question (i.e., they presume evolved individuality as a starting point for evolutionary explanations). It is a mistake to think we can get a full Darwinian picture by only considering natural selection with its standard presuppositions. Ellen Clarke writes, "we are obliged to explain the appearance of higher-level organisms, rather than presuppose them as targets in explanations of how selection can act at a higher level."[75] But this appearance of individuals is a mystery. Maynard Smith and Szathmáry pose a similar question in their seminal work on the subject: "Why did not natural selection, acting on entities at the lower level, disrupt integration at the higher level?"[76] Consider the insufficiency of Dawkins and Hulls' concepts of replicators and interactors (or for Dawkins, "vehicles"). For the latter to be defined as an entity that interacts with its environment as a *whole*, we must presume wholeness, that is, individuality. As to the former, the notion of high-fidelity replicators copying themselves and persisting eternally presumes a capacity for high-fidelity replication. Individuality and copying fidelity are both evolved characters, and so Okasha notes, "we cannot build these notions into the very concepts used to describe natural selection."[77]

Suppose you say that what makes something a unified whole, as opposed to some loose collections of things, is that the parts interact.[78] You might then add the additional requirement that lower-level individual actions need to be fitness affecting interactions, such that they have a "shared evolutionary fate."[79] But then you might suggest that while interaction is necessary, to be considered an individual requires sufficient functional integration and coordination (e.g. the collective must have evolved policing mechanisms or a division of labor). [80] But here is where we encounter a chicken-egg type of mystery: functional integration is a group-level adaptation. To evolve, e.g., policing mechanisms, is for selection to favor more refined (individualized) structures of collective entities. Therefore, you couldn't require functional integration as a necessary condition for collective level individuality if collective level individuality is required to be a candidate for such evolutionary refinement! Yet, it is not easy to avoid citing various biological mechanisms as

[75] Ellen Clarke, "Origins of evolutionary transitions," *Journal of biosciences* 39, no. 2 (2014): 303-317.
[76] Maynard Smith & Szathmáry, *The Major Transitions in Evolution,* 8.
[77] See Chapter 1 of Okasha, *Evolution and the levels of selection.*
[78] Daniel W. McShea, "The hierarchical structure of organisms: a scale and documentation of a trend in the maximum," *Paleobiology* 27, no. 2 (2001): 405-423.
[79] Sober and Wilson, *Unto Others*.
[80] Richard Michod and Aurora Nedelcu, "On the reorganization of fitness during evolutionary transitions in individuality," *Integrative and Comparative Biology* 43, no. 1 (2003): 64-73.

somehow both the result of, and prerequisites for, higher-level selection in our explanations of transitions.[81]

We are left with a seeming paradox that group features have to be selected before the group exists (*qua* group, not a mere collection). We can draw a comparison to similar mysteries regarding agents that seemingly bootstrap themselves into existence. This idea was raised in section 2.3 and I will attend to it again in chapter 6. I will further suggest that the two literatures might be able to help each other with novel approaches to these problems (e.g. the problem of incentive compatibility between individuals and group agents, with the biological notion of fitness alignment (or convergence) or "shared fate" preceding an evolutionary transition from aggregates to new wholes).

3.4 A Closer Look – Selected Issues and Accounts

In this section I will first give an overview of two leading accounts of biological individuality: Ellen Clarke's evolutionary account and Thomas Pradeu's immunological account. These are two of the most prominent accounts which nicely portray quite different points on the spectrum of views on biological individuality and will inform later discussions. Following this, I will feature the work of Maureen O'Malley, who distinguishes "units of selection" and "units of viability" while emphasizing the need to recognize the metabolic process of individuals as a collaborative, multi-species affair. This distinction (which maps onto the previously introduced distinction between evolutionary and physiological individuality in section 3.2) paves the way for a pluralistic approach to biological individuality, which would allow us to retain the insightful aspects of different accounts (like those of Clarke and Pradeu). From there, I will go onto discuss some interesting cases that put pressure on our ways of thinking about biological individuality and further motivate the case for pluralism.

3.4.1 Evolutionary Individuality

[81] For a more nuanced treatment of the chicken-egg mystery, see Clarke, "Origins of evolutionary transitions," 310.

Over the last decade, Ellen Clarke has developed and refined an evolutionary account of biological individuality that comports well with the literature on evolutionary transitions in individuality. Clarke's account is predicated on the following. Biology concerns the living, and the cohesiveness of living things is an evolved character. Only evolution can explain why we find life, not as "some sort of homogenous soup," but arising in "cohesive packages… so that it can participate in selection processes at various hierarchical scales" ("so that" not, of course, offered in some teleological sense).[82]

An "evolutionary individual" then, is one that is capable of undergoing natural selection.[83] So for Clarke, the capacity to undergo selection is a necessary condition of biological individuality. If we can identify mechanisms which preserve, influence or otherwise account for this capacity, we can locate individuals at the hierarchical level that these mechanisms are situated. Clarke refers to this as her "levels-of-selection" definition of biological individuality. Suggesting that some conceptions of biological individuality will be more useful than others, Clarke sees greatest utility in the evolutionary conception of individuality, specifically in its power to predict and explain evolutionary processes.[84]

Clarke argues that competing individuality criteria can be unified if these criteria can be evaluated with respect to how they "Darwinize" the purported individuals, that is, make the purported individual more or less capable of undergoing natural selection at the level of that purported individual. So then, Clarke's evolutionary account is a monist, functionalist account. She argues that biological individuality can be multiply realized.[85] She defines biological individuals as "all and only those objects that possess" both of two kinds of individuating mechanisms: policing and demarcating mechanisms.[86] Policing mechanisms inhibit *within-object* selection, while demarcating mechanisms increase (or maintain) the capacity for *between-object* selection.[87] (I will give some brief examples below). Thus, the

[82] Ellen Clarke, "Is Evolution Fundamental When It Comes to Defining Biological Ontology? Yes," in *Current Controversies in Philosophy of Science*, eds., Shamik Dasgupta, Ravit Dotan, and Brad Weslake, 104-118 (New York: Routledge, 2020).
[83] Though, unlike others (e.g. O'Malley), Clarke gives less attention to the non-living entities included in this category.
[84] Consistent with what I have introduced above, Clarke distinguishes evolutionary concepts from organizational concepts, importantly noting that they are not "rival definitions of a single concept." Clarke, "Is evolution fundamental," 105.
[85] Clarke, "The multiple realizability of biological individuals."
[86] Clarke, "The multiple realizability of biological individuals," 424.
[87] Or "*positively grounds* the capacity of an object to participate in a between object selection process" (emphasis added). Clarke, "The multiple realizability of biological individuals," 433.

phrase "individuating mechanism" is more general, and any given individuating mechanism can be described as either a policing or demarcating mechanism. These mechanisms can function with greater or lesser individualizing efficacy, which means individuality is a property that obtains to a degree.

So biological individuality requires at least one of each mechanism. But each mechanism can be realized in a number of ways. An example of policing would be an immune system targeting defective mutant cells that threaten to multiply at cross-purposes to the multi-cellular entity as a whole. A similar example would be a worker honeybee eating the eggs of a "rogue" sister bee (as on the colony level, the queen's offspring function similarly to that of a germ line in a multicellular organism). We can see the latter most clearly as an example of eliminating competition between variants. Natural selection requires variation and so when one variant eliminates others, this can reduce evolutionary competition. We can find similar examples for demarcation mechanisms, one example being that of a cell wall or other type of boundary (e.g. skin), which limits mixing and migration of the constituent parts. Demarcation increases "between-*collection* selection," i.e. the mechanism which has a role in cementing the individuality at a higher level. Another example, according to Clarke, is sex: "by increasing between-organism genetic variance, sexual recombination boosts or enhances selection at the between-organism level." This can be contrasted with the aforementioned immunological policing mechanisms that guard against such variation, reducing "selectability" within the organism.[88]

Though Clarke's account is monist, it can accommodate many of the particular candidate criteria that had previously been offered as competing requirements based on their *capacity to individualize* (e.g. germ/soma separation, bottle neck, genetic homogeneity). Clarke is arguably able to accommodate many of the candidate criteria (e.g. as documented by Lidgard and Nyhart) and offers her account as a solution to the problem that has historically plagued accounts of biological individuality. For the individuality debate, it is this aspect of Clarke's account that is most salient.

Recall the common view that biological individuality admits of degrees, with individuality expressed more or less at higher or lower levels. Clarke does note the possibility of

[88] Clarke, "The multiple realizability of biological individuals," 424.

articulating an exclusive principle (i.e. capacity to respond to selection at only one level).[89] However, Clarke leaves room for the messiness of multi-level selection, admitting of borderline cases of individuality-in-progress. In other words, Clarke is not endorsing an exclusionary (i.e. "anti-nesting") view, in which these mechanisms specify a single hierarchical level on which individuality will obtain. She therefore invokes the language of dominance and determination instead of mutual exclusion: "the two classes of mechanisms determine which of two opposing hierarchical levels of selection will dominate, and so determine facts such as whether altruistic traits can spread through the population."[90]

These aspects of Clarke's account are ripe for explorations between agency concepts and biological individuality. Questions for future work include: how might Clarke's thoughts on individuality inform Rovane's contrasting account of agency or vice versa (recall that Rovane holds an exclusive or anti-nesting view of agency)? Such discussions will be relevant to the subjects of incentive compatibility and inclusive versus exclusive agency (see Chapter 6).

More generally, a key notion I take from Clarke's account is the *maintenance* of individuality. For Clarke, what defines an individual is the presence of mechanisms able to constrain and maintain processes which ensure unity. For Clarke, what this unity ensures is the capacity to undergo selection (and thus count as an individual on an evolutionary account). However, the unity that these mechanisms preserve is often itself seen as the individuating factor (this can be seen in other accounts which focus on organization and physiology). We should not therefore assume that fruitful connections between biological individuality and group agency must make reference to the evolutionary aspects of biological individuality. If all creatures were created by God, we could still meaningfully explore these connections. Having said that, evolutionary considerations will largely figure in chapter 4, in which I suggest that a particular understanding of "rational unity" can justifiably serve as a common conceptual currency between the two bodies of literature examined in this thesis.

Before moving on, I want to point out an important aspect of multi-level selection theory and evolutionary accounts of individuality. The status of an individual *qua* individual does not depend on its "place" in the hierarchy. Samir Okasha, who proposes a "rank free" approach

[89] See also Peter Godfrey-Smith, "Darwinian Individuals," in Bouchard and Huneman, *From groups to individuals*, 26.
[90] Clarke, "The multiple realizability of biological individuals," 432.

to evolutionary transitions, makes this point well. Okasha's own work on multi-level selection theory uses the language of "particles" and "collectives,"[91] though we might otherwise say "individuals" and "groups." Okasha's emphasis on the abstract reminds us that "all entities in the hierarchy, apart from those at the very bottom, are simultaneously groups and individuals, depending on our choice of focal level."[92] Therefore, "organism" should not denote a rank in the ecological hierarchy, just as the old Linnaean categories of "family" or "order" should hold no rank for any given monophyletic taxon. Okasha is not saying the organism concept is a meaningless concept. On the contrary, what counts as an organism is that which meets the relevant criteria (e.g. "exhibits sufficient functional integration").[93] But to classify an organism is not to denote rank within an ecological hierarchy. To meet the criteria of individuality just means *to belong to the hierarchy*.[94] There are single-celled organisms, and multicellular organisms. They are organisms not by virtue of rank (or degrees of nestedness or size or whatever), but because they met the criteria for individuality (e.g. functional integration, low conflict, etc.).

The question of whether or not rank or status is conferred due to degrees of nestedness, or whether expanding or shrinking the hierarchy comes at the cost of one's status, will make for an interesting comparison with social issues regarding agents and persons. In section 5.1 we will confront problems List and Pettit face in accommodating both group agency and normative individualism (concluding that it is not an individual human's place in the hierarchy that confers protected status, but rather their intrinsic nature, contrary to LP's explicit departure from "instrinsicism"). In chapter 6, I will briefly sketch for future work the rival views of inclusive versus exclusive agency (in which the latter would not allow for nested agency even in principle).

3.4.2 Physiological Individuality

[91] Okasha, *Levels of Selection.*
[92] Samir Okasha, "Biological Ontology and Hierarchical Organization: A Defense of Rank Freedom," in *The Major Transitions in Evolution Revisited*, eds., Brett Calcott and Kim Sterelny (Cambridge.: MIT Press, 2011), 60.
[93] Okasha, "Biological Ontology and Hierarchical Organization," 58.
[94]Note that Okasha is using "organism" and "individual" interchangeably here picking out that which participates in natural selection (as multi-level theorists will acknowledge that to some degree, selection acts on e.g. colonies and the organisms that make them up, as well as the cells that make up the organisms).

Some accounts collapse any distinction between "organism" and "individual."[95] Others see organisms as paradigm biological individuals. Often the latter classification is done on an informal and intuitive basis, favoring animals that more closely resemble humans. Others might take a formal approach, saying that what counts as a paradigm organism will be that which meets a number of conditions, possess qualities, or qualifies as more than one *kind* of individual. There is, at the least, a growing agreement that it is unhelpful to use "individual" and "organism" interchangeably.[96]

Pradeu's Immunological Account

Thomas Pradeu offers an immunity-based account of biological individuality.[97] "An immunological individual is a functionally-integrated whole made up of heterogeneous constituents that are locally interconnected by strong biochemical interactions and controlled by systemic immune interactions." [98] The immune system, Pradeu argues, has a role to play in showing how individuals are delineated, how they are unique, and how they can be considered the same over time. One's immunological record is like an ever-updated thumbprint – a history of one's encounters with and within the environment. Consider that even monozygotic twins will have different immune systems such that they can be uniquely identified.[99]

One particular strength of Pradeu's argument is that it can handle multi-lineage or otherwise heterogeneously composed organisms. Human immune systems, for example, "know" to attack foreign pathogens but also "recognize" and pass over "friendly" gut bacteria (an example of how immune systems delineate individuals). Thus, one of Pradeu's main contributions is to make us aware that unlike previous conceptions of immunology, the immune system does not simply accept endogenous and reject exogenous materials (with "self" and "not self" applying respectively).[100] For Pradeu, "any entity which interacts

[95] E.g., Pepper & Herron, "Does biology need an organism concept?"; Queller and Strassman, "Beyond Society"; Clarke, "The multiple realizability of biological individuals."
[96] e.g. Thomas Pradeu, "Organisms or biological individuals?"
[97] Thomas Pradeu, "Immunity and the emergence of individuality," in Bouchard and Huneman, *From groups to individuals*, 77-96.
[98] Pradeu, "Organisms or biological individuals?" 84.
[99] Pradeu, *The limits of the self.*
[100] Pradeu, *The limits of the self.*

regularly with the immune system and is not eliminated by it is part of the physiological individual."[101]

Pradeu argues that only his conception can account for "the unity of the organism" in light of its heterogeneous constitution. Unlike many others, Pradeu takes "organism" to be a "pertinent scientific category," not merely a folk term. Pradeu defines organisms immunologically, then shows which sorts of things have immune systems "in the sense of a system of biochemically specific interactions leading to the rejection of some living entities."[102] Understood as such, the term "organism" should not be limited to multi-cellular organisms, for Pradeu argues that even prokaryotes have a kind of immune system (e.g. CRISPER-Cas systems). An organism is defined by its organization - its functional integration of interdependent sub-systems and component parts. Pradeu suggests that an immune system is what achieves this organization. Therefore, for Pradeu, the term "organism" should refer to a subset of biological individuals that are "cohesively organized through an immune system."[103]

To be clear, the immune system does not *equal* the organism. The immune system "defines what will be accepted and what will be rejected by the organism."[104] In other words, the "distinction between the entities which will stick together as constituents of the organism and those which will be rejected from the organism, is made by the immune system."[105] Pradeu does not doubt other physiological perspectives might also offer an account of organismic individuality, and is thus not necessarily at odds with those who would instead prefer to say organisms qualify as such by virtue of their functional, developmental, or metabolic capacities (and we will explore more of this terrain via the work of O'Malley below). The immune system, on Pradeu's account, is a useful way of successfully identifying organisms.[106]

---

[101] Pradeu, "Organisms or biological individuals?" 805.
[102] Pradeu, "Immunity and the emergence of individuality," 81.
[103] Pradeu, "Immunity and the emergence of individuality," 89.
[104] Thomas Pradeu "What Is An Organism? An Immunological Answer," *History and Philosophy of the Life Sciences* 32, no. 2/3 (2010): 253.
[105] Pradeu, "What Is An Organism?" 253.
[106] I think this aspect of Pradeu's account raises questions, e.g., about "germ free" mice used in laboratories, immunocompromised "bubble boys," and microbes with and without CRISPR-Cas. Around 40% of bacteria and 90% of archaea possess CRISPR Cas. Consider two bacteria, one with CRISPER-Cas and one without. However similar these bacteria may be in other ways, we might, on Pradeu's account, be forced to say that one is an organism and the other is not. Where any bacteria and archaea are found to not have CRISPR Cas, they must, to count as an organism, have some other kind of immune system; or, many bacteria and archaea are not organisms, while otherwise similar bacteria and archaea are organisms. The latter is highly counter-intuitive.

This immunological perspective then is consistent with Clarke's emphasis on multi-level selection. Pradeu makes a distinction between "ex ante mechanisms" and "ex post" mechanisms. That is, some mechanisms prevent conflict, while others remedy conflict. An example of an *ex ante* mechanism is the evolution of the germ-soma divide,[107] which preserves homogeneity by virtue of the fact that soma cells are evolutionary dead ends - their interest is tied to the organism as a whole, through which their genotype can only be passed on via the germ line. An example of an *ex post* mechanism would be a process triggered by an animal's lymphocytes in which the immune system has evolved to recognize, target, and eliminate tumors which regularly arise (e.g. according to the "immune surveillance hypothesis").[108] The connection to Clarke's evolutionary account is notable. Pradeu sees an organism's immune system as performing both demarcating and policing functions, to use Clarke's preferred terminology. Put simply, an immune system rejects what does not belong, foreign and domestic, and can thus be seen as establishing the boundaries of an individual as well as policing internal mutants.

However, Pradeu's account does not merely fold into Clarke's, with the immune system serving as one of many potential individuating mechanisms. This is because for Clarke, the higher-level individuality that sub-system mechanisms maintain is still *evolutionary* individuality. In a classic case of "in between" individuality, the slime mold has an immune system with sentinel cells that cruise around the grex (or "slug") and stick to bacterial pathogens. Thus, the immune system maintains the higher-level individuality of the slime mold. But we need not commit to thinking of this higher-level individual as an *evolutionary* individual. Should we decide there is not adequate selection between slime molds but rather only between the amebae that make them up, we can still meaningfully consider the slime mold an individual organism.

In later work, Pradeu articulates his view in different terms with less emphasis on the "organism," preferring to focus on the distinction between "physiological individuals" and

---

For percentages see Grissa et al., "The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats," *BMC bioinformatics* 8, no. 1 (2007): 1-10.

[107] Pradeu himself thinks the "role of the germ-soma separation in biological individuality is much exaggerated." Pradeu, Immunity and the emergence of individuality," 86.

[108] Frank Burnet, *Immunological surveillance* (Oxford: Pergamon Press; 1970).

"evolutionary individuals."[109] This is desirable so as to not further cloud the issue on what counts as an "organism." Even in the philosophical work which contents itself to talk about organisms descriptively, as a folk concept, confusion may arise. We can compare Pradeu's work with that of O'Malley and Godfrey-Smith (to be discussed subsequently). At first blush, it seems that Pradeu's distinctions are similar. For example, we can note that a gene is one example of something which is a biological individual yet not an organism; it is subject to natural selection and is therefore an evolutionary (or Darwinian) individual. But the *reason* it is not an organism will be different. For Pradeu, a gene will not be on organism if it lacks an immune system. For O'Malley and Godfrey-Smith a gene is not an organism because it lacks metabolic sufficiency; only evolutionary individuals which can persist and reproduce "with their own machinery" further qualify as organisms. For these latter philosophers, "organism" is the term we intuitively apply to the subset of biological organisms that are both paradigm persisters and units of selection. In other words, the set of Pradeu's "organisms" would overlap with O'Malley and Godfrey-Smith's. But even if they did not, the distinction of criteria is non-trivial. For example, Pradeu claims "all living things have immune systems" and so while some would say chromosomes are not individuals because they are not living, this would only follow indirectly for Pradeu. [110]

In conclusion, Pradeu's account satisfyingly integrates the multi-selection perspective with both our modern knowledge of multi-lineage symbiosis and our intuitions about organisms.

O'Malley's Metabolic Account

Some will subscribe to the distinction between evolutionary and physiological senses of the term "individual," yet have a different story about how the latter achieve the organization that warrants our attribution of individuality. Of particular interest to O'Malley (in collaboration with Dupre and, separately, Powell) are types of purported biological individuals which are neither Darwinian individuals nor what we typically think of as organisms (in the paradigmatic sense), yet are still "metabolic wholes."[111] Such metabolic individuals[112] are

[109] Pradeu, "Organisms or biological individuals?"
[110] Pradeu, "Organisms or biological individuals?" Cf. those who consider many parts of individuals to be living e.g. Robert Wilson, *Genes and the agents of life."*
[111] Maureen O'Malley, "Is Evolution Fundamental When It Comes to Defining Biological Ontology? No," in Dasgupta et al., *Current Controversies in Philosophy of Science*, 95.
[112] This seems a natural term to use here though authors have variously opted to speak of "metabolic wholes" or "metabolic persisters."

consistent with Godfrey-Smith's and Subrena Smith's characterization of "persisters," defined as such with no further criteria e.g. reproductive capacity, belonging to a lineage.[113] The kind of entities O'Malley has in mind here are symbionts and other cooperative ventures of the living world: "persistent and regularly regenerated metabolic wholes without a common inheritance – composed of sub-entities that have separable biological and evolutionary individuality."[114] The phrase "separable biological evolutionary individuality" is not intended to be at odds with the aforementioned categorical scheme, rather it shows that two symbionts are of a kind intuitively viewed as a biological individual in their own right, i.e. organisms, apart from the metabolic collaboration.

O'Malley makes clear the different classificatory and explanatory distinctions of these different concepts:

> [B]iological individuals are the units of ecological interaction and survival, achieved most basically by their metabolic capabilities. The explanatory focus for such entities is their persistence in and impact on the world. A Darwinian individual, on the other hand, is a unit of selection regardless of whether it can maintain itself metabolically. The explanatory focus for these entities is fitness measured as multiplication of the entities… we explain biological individuals by viability, and Darwinian individuals by fecundity.[115]

In other words, we may have *units of viability* composed of two or more *units of selection*. Having surveyed both Clarke and Pradeu's account, we can now see what has motivated those in the recent literature to accept a framework that makes use of both evolutionary and physiological conceptions of individuality. O'Malley's added emphasis on the ubiquitous multigenomic collaboration among living things partially motivates what I will argue in section 5.2: useful agential explanations are available to us independent of the contingency and complexity of our various attributions of individuality.

[113] Peter Godfrey-Smith, *Philosophy of biology* (Princeton University Press, 2013), 77.
[114] O'Malley, "Is Evolution Fundamental? No," 95.
[115] O'Malley, "Is Evolution Fundamental? No," 97.

To give one example of O'Malley's "co-metabolizing consortia:"[116] Endosymbiotic rhizobial bacteria provide nitrogen to legume plants which depend on it for growth. The bacteria likewise benefit (though not essentially) by achieving greater reproductive output in the plant's root nodule than it would in the soil.[117]

In her work with Dupre, O'Malley stresses the relevance of *collaboration* in life's achievement of individuality.[118] The notion of "selfishness" is widespread in biology, and if over-emphasized could easily eclipse the fact that biological individuality emerges evolutionarily from a consortium of individuals. Dupre and O'Malley suggest that "competitive activity is a transitional rather than terminal state."[119] Dawkins' notion of the "selfish gene" for example, belies the fact that "molecular replication… has always been…. the achievement of ensembles of molecules."[120] Both cooperation and competition, they argue, should both be considered within the broader notion of what they call "collaboration," to be understood as "interactions between components of a system that lead to different degrees of stability, maintenance or transformation of that system."[121]

This contrast of perspective may mark a turning of the tide and speaks further to how at odds our concepts of individuality are with the origins of individuality. "The unit of selection, the entity in which selfishness may perhaps be expected as the norm, is a collaboration of many different lineage-forming entities."[122] We can look to the history of the evolution of the eukaryotic cell and see evidence of such collaborations. As previously mentioned in section 3.3.1, mitochondria are organelles within eukaryotic cells, but were once free-living bacteria and still reproduce by their own genetic mechanisms. Mitochondria could not survive outside of the cell, nor could a cell survive without the essential work performed by the

[116] O'Malley, "Is Evolution Fundamental? No," 95.
[117] R. Ford Denison and E. Toby Kiers, "Life histories of symbiotic rhizobia and mycorrhizal fungi," *Current Biology* 21, no. 18 (2011): R775-R785. For a similar but stronger case (arguing for evolutionary individuality) concerning symbiotic fungus attached to plant roots, see Daniel J. Molter, "On mycorrhizal individuality," *Biology & Philosophy* 34, no. 5 (2019): 1-16.
[118] John Dupré and Maureen O'Malley, "Varieties of Living Things: Life at the Intersection of Lineage and Metabolism," *Philosophy, Theory, and Practice in Biology* 1 (2009): 1-25.
[119] Dupré and O'Malley, "Varieties of Living Things," 14. One could argue this is perspectival and subject to reversal (an egg is a chicken's way of making another chicken; a chicken is an egg's way of making another egg).
[120] Dupré and O'Malley, "Varieties of Living Things," 15. A most interesting (though implicit) suggestion that shared concepts could span biology and chemistry, such that in future work we may try to extend "individuating" mechanisms to be cashed out in chemical and physical terms.
[121] Dupré and O'Malley, "Varieties of Living Things," 2.
[122] Dupré and O'Malley, "Varieties of Living Things," 14.

mitochondria. Or consider *buchnera aphidicola*, bacteria that live within, are dependent on, and produce amino acids for, aphids. We call them endosymbionts but perhaps, because of their dependence on host cells, we should see them as having a similar status to organelles.[123] This continuum from free living bacteria to endosymbiont to organelle is demonstrative of the evolution of life and individuality resulting from collaboration across hierarchical levels.

Intracellular processes, however, are not only cooperative as "competitive reproductive relationships may also exist between organelles or plasmids," hence Dupre and O'Malley's suggestion that such varied dynamics be seen in terms of the over-arching notion of "collaboration."[124] They take "functional wholeness" to be the "basis of any attribution of autonomy." Functional wholeness is a "characteristic of collaborative interactions, almost always involving diverse entities."[125] They make the further very interesting assertion that biological individuals aren't living so long as they are individual: "single animal or plant cells are only truly alive when they are collaborating with other cells."[126]

Biological individuals are persisters with ancestry. We can conceive of the living, as articulated by Dupre and O'Malley, as "lineage-forming, metabolically collaborative matter."[127] We can think of living individuals as possessing both a code and the means of executing that code. The means of persisting and executing the code are achieved *via* metabolic processes, internal and external. But "*lineage-exclusive* autonomous reproduction" is a myth (emphasis added). Reproduction requires metabolic processes, and therefore indirectly requires symbiotic microbes which fulfil metabolic functions.[128] "All organisms, when conceived as the functional wholes that interact with their surroundings, are multi-lineal and multigenomic" (the latter implying "codes" plural).[129]

[123] Jan O. Andersson, "Evolutionary genomics: Is Buchnera a bacterium or an organelle?" *Current Biology* 10, no. 23 (2000): R866-R868.
[124] Dupré and O'Malley, "Varieties of Living Things," 10-11. See Clarke's comments on plant mutations ("trying out" different phenotypes to adapt to the environment). Ellen Clarke, "Plant individuality: a solution to the demographer's dilemma," *Biology & Philosophy* 27, no. 3 (2012): 321-361.
[125] Dupré and O'Malley, "Varieties of Living Things," 11.
[126] Dupré and O'Malley, "Varieties of Living Things," 10. In the spirit of this thesis, I cannot help also wondering, are there any agents without inter-agency?
[127] Dupré and O'Malley, "Varieties of Living Things," 16.
[128] Dupré and O'Malley, "Varieties of Living Things," 13.
[129] Dupré and O'Malley, "Varieties of Living Things," 12.

Earlier I said that endorsing a set of individuality criteria is often an advocation for what is important. O'Malley makes the case for the importance of organizational criteria to contrast with Clarke's view about evolutionary importance. As we have seen, those like Pradeu are also interested in heterogeneously constituted multi-species collaborations, some of which are sometimes called "holobionts." We will turn now to these and other cases in which the status of individuality is up for debate.

3.5 Curious Cases: Holobionts, Super-Organisms, Extended Individuals

Clarke's account collapses any distinction between so-called "organisms" and individuals, arguing that evolutionary individuals, understood as multiply realized on her monist account, are the referent of our questions about real biological individuals. Pradeu's immunological account shows one way of being a physiological individual (or for some, an "organism"), whether or not that physiological individual is also an evolutionary individual.[130] For Pradeu, individuals are delineated by the continuity of specific biochemical interactions (irrespective of endogenous origin). For others, such as O'Malley and Godfrey-Smith (featured in section 3.6.1), physiological individuals are metabolic persisters, functionally integrated with co-dependent components (whether or not they have an immune system or capacity to undergo selection at their own level).[131] The following three cases are exemplars of the kinds of entities about which philosophers of biology are sometimes puzzled. Together with some of the considerations we have weighed so far, what is presented in this section aims to demonstrate aspects of the debate that motivate the adoption of individuality pluralism by some philosophers of biology.

3.5.1 Holobionts

The status of some purported individuals is up for debate. Some chunks of the biological world exhibit high degrees of functionally integrated unity despite their heterogeneous composition. In section 3.4.2, we were introduced to the notion of metabolic individuals –

[130] Re: "or organism:" Recall that the usage of "organism" varies in the literature. Unhelpfully it sometimes picks out a paradigm individual which is both a physiological and evolutionary individual, other times it refers simply to a physiological individual (though that physiological individual will be comprised of evolutionary individuals).

[131] "Co-dependent" should not mean equally or exclusively co-dependent. A plant may crucially depend on the activities of certain microbes, but the microbes themselves may find the benefits derived from interaction with the plant elsewhere in a "free living" state (i.e. apart from the symbiotic coupling).

multispecies collaborations unified by e.g., tight persistent nutrient exchange. Yet modern biology (e.g., the study of hereditary symbiosis) suggests that even *paradigm* individuals such as mammals are also multispecies groups, and always have been.[132]

Lynn Margulis coined the term "holobiont" to pick out ecological units of tight collaboration between multi-cellular hosts and their endosymbiotic microbiota (bacteria, viruses, fungi).[133] The most common example is perhaps a human being and its microbiome (the microbes in a human's gut, on their skin, in their mouth, etc.). Thus, holobionts are multi-species entities. What we make of "entity" is a source of debate.

Some argue that holobionts are ecological communities[134] while others argue they are biological individuals.[135] Of the latter, various accounts will specify the kind of individual that holobionts are taken to be, e.g., developmental,[136] immunological,[137] or metabolic.[138] Some aspects or types of individuality are often emphasized over others, such as whether holobionts are units of selection[139] or "interactors."[140] Suarez and Stencel argue that holobionts are both individuals and ecological communities (qualifying as one or the other depending on whether we privilege the "perspective" of the host or the microbes).[141] Lloyd and Wade provide more nuanced categories of holobionts (e.g. "euholobiont" and "demibiont"), more sensitive to the different kinds of symbiosis (i.e. parasitism,

[132]Jan Sapp, "Paul Buchner (1886–1978) and hereditary symbiosis in insects," *International Microbiology* 5, no. 3 (2002): 145-150.
[133] Lynn Margulis, "Words as battle cries: symbiogenesis and the new field of endocytobiology," *Bioscience* 40, no. 9 (1990): 673-677.
[134] E.g. Nancy A. Moran and Daniel B. Sloan, "The hologenome concept: helpful or hollow?" *PLoS biology* 13, no. 12 (2015): e1002311.
[135] E.g. Llana Zilber-Rosenberg and Eugene Rosenberg, "Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution," *FEMS microbiology reviews* 32, no. 5 (2008): 723-735.
[136] James Griesemer, "Reproduction in complex life cycles: toward a developmental reaction norms perspective," *Philosophy of Science* 83, no. 5 (2016): 803-815.
[137] Lynn Chiu, and Gérard Eberl, "Microorganisms as scaffolds of host individuality: an eco-immunity account of the holobiont," *Biology & Philosophy* 31, no. 6 (2016): 819-837. More precisely, the authors offer for a hybrid "eco-immunological" view for which you might say holobionts' physical individuality is overdetermined.
[138] E.g. Dupré and O'Malley, "Varieties of Living Things"; Smith, "Organisms as persisters"; Gilbert et al., "A symbiotic view of life."
[139] E.g. Feldhaar, Heike, "Bacterial symbionts as mediators of ecologically important traits of insect hosts," *Ecological Entomology* 36, no. 5 (2011): 533-543. "[T]he host including all symbionts, should be regarded as the unit of selection as the association between host and symbionts may affect the fitness of the holobiont depending on the environment" (533).
[140] Derek Skillings, "Holobionts and the ecology of organisms: multi-species communities or integrated individuals?" *Biology & Philosophy* 31, no. 6 (2016): 875-892; Elizabeth Lloyd, "Holobionts as units of selection: holobionts as interactors, reproducers, and manifestors of adaptation," In *Landscapes of Collectivity in the Life Sciences,* eds. S. B. Gissis, E. Lamm and A. Shavit, (London: MIT Press, London, 2017): 351– 368.
[141] Suárez and Stencel, "A part-dependent account of biological individuality."

commensalism, mutualism) and the degree to which the host and microbes have co-evolved and co-adapted.[142]

On another view, holobionts may be an example of physiological individuals which are neither evolutionary individuals nor organisms (e.g. but rather ecological communities).[143] Arguably then, Pradeu's "physiological individuals" or "organisms" (in his sense) are holobionts.[144] The upshot is that physiological individuals can be a larger, more encompassing term. "Physiological individual" gives us a natural, if imprecise, way to meaningfully call e.g. human beings together with their microsymbionts "individuals" without unintuitively abstracting away bits of human cells from other symbiotic cells (e.g. gut bacteria) of exogenous origin.

3.5.2 Super Organisms

"Super-organism" most generally refers to organisms that are comprised of other organisms. At first glance, this category seems of high relevance to this thesis topic (intuitively paralleling "group agent"). But like so many things, there is disagreement about its meaning and ambiguity in its usage. Jack Wilson observes two common ways the term "superorganism" is used. Sometimes it is used to denote a collection of integrated living entities that, for whatever reason, we feel don't quite make the grade of organismal status. Other times the term "superorganism" is used to describe a collection of organisms that together "act as a unit of selection."[145]

In a classic paper on superorganisms, Wilson and Sober argue that "groups and communities can be organisms in the same sense that individuals are."[146] Why not super-*individual*? They suggest that the term "organism" is best for focusing on units of functional organization. It is clear for Wilson and Sober that superorganisms exist in nature – they are real. Some groups undergo natural selection just like individuals, becoming functionally organized just as

[142] Elisabeth Lloyd and Michael J. Wade, "Criteria for holobionts from community genetics," *Biological Theory* 14, no. 3 (2019): 151-170.
[143] Skillings, "Holobionts and the ecology of organisms: multi-species communities or integrated individuals?"
[144] Wilson and Barker, "Biological Individuals."
[145] Jack Wilson, "Ontological butchery: organism concepts and biological generalizations," *Philosophy of Science* 67 (2000): S301-S311.
[146] David Sloan Wilson and Elliott Sober, "Reviving the superorganism," *Journal of theoretical Biology* 136, no. 3 (1989): 337.

organisms do.[147] They suggest that seemingly competing concepts are unnecessarily ambiguous and a convergent scientific articulation of a theory of superorganisms is not only possible, but needed. Wilson and Sober define a superorganism as "a collection of single creatures that together possess the functional organization implicit in the formal definition of an organism."[148]

As should be clear from the individuality debate itself, there is no consensus on what it takes to be an organism, and so Wilson and Sober's view may be subject to criticism depending on what account of organismality one adopts. We might also wonder about the exact analogy we are meant to make, as they themselves seem reluctant to make precise parallels: "Individuals acquire the status of alleles and the groups acquire the status of superorganisms… Individuals can properly be viewed as *genes/organs* within a superorganism."[149] They are perfectly explicit, however, in acknowledging that the analogy is only possible from the perspective of multi-level selection theory, which they endorse.

With this latter point we see Wilson and Sober distancing themselves from the tradition that rejects multi-level selection theory– those that maintain a position of explaining all evolutionary change based on the effects of genes, e.g. selfish-gene theory. The title of their famous paper is "*Reviving* the Superorganism," because for a time the idea of group-level selection was scorned in the field.[150] It was thought, for example, that Hamilton's notion of inclusive fitness could sufficiently explain sterile castes in social insects.[151] Wilson and Sober insist that inclusive fitness is not a competing theory, but rather a different way of explaining multi-level selection (this will feature in an example of level-shifting in section 5.2). This point is made clear with the following example:

> Consider a mutant behavior expressed in sterile workers that increases the fecundity of the queen. The behavior obviously increases the inclusive fitness of the worker, but the frequency of the allele does not increase within the colony. It evolves only by

[147] "Natural selection endows [some] groups with the same properties of functional organization that we normally associate with individual organisms." Wilson and Sober, "Reviving the superorganism," 342.
[148] Wilson and Sober, "Reviving the superorganism." 339.
[149] Wilson and Sober, "Reviving the superorganism," 341. Emphasis added.
[150] Emphasis added.
[151] William D. Hamilton, "Innate social aptitudes of man: an approach from evolutionary genetics," *Biosocial anthropology* 53 (1975): 133-155.

> between-colony selection; colonies bearing the allele out-produce colonies that do not.[152]

The upshot is that the maximization of inclusive fitness and group level selection are not two separate processes. Inclusive fitness theory grants us an insightful perspective - but make no mistake: it *requires* between-colony selection.

I mention both the sociological point and the tidbit from the group-selection debate to show that although Wilson and Sober's definition would seemingly put them in the "functional integration" camp of biological individuality, their position is firmly rooted in evolutionary biology, with superorganisms forged *via* selection pressures operating between (rather than among) groups. When the latter is emphasized, it is often dubbed an "evolutionary account," offering criteria for "evolutionary individuality," which may or may not coincide with other types of individuality (of which there are a range on offer).

It is a side effect of the fact that historically the terms "organism" and "individual" have been used interchangeably that those wanting to refer to unified biological groups resorted to the qualifying and superlative title, "superorganism."[153] Jack Wilson suggests that the term is

> sometimes used to describe the functional organization of entities composed of more than one organism. The term was necessary because, given the ambiguous nature of 'organism,' there was no term that unambiguously referred only to the functional organization of a living entity independent of other properties such as the origins of the entity.[154]

It is for this reason that Jack Wilson advocates for what he feels is a more appropriate, alternative title, "functional individual" (although others with different categories might hasten to add that superorganisms like ant colonies can also be evolutionary individuals, i.e. units of section, which is Haber's point below).

---

[152] Wilson and Sober, "Reviving the Superorganism," 346.
[153] David Hull, "Are species really individuals?" *Systematic zoology* 25, no. 2 (1976): 175.
[154] Jack Wilson, *Biological individuality: the identity and persistence of living entities* (Cambridge University Press, 1999), 63.

On Jack Wilson's account (explored more fully in section 3.6.2), a "functional individual" is one that is not determined by its genetics, developmental history, nor even the history of its component parts, but rather the causal relations between parts. Yet Wilson confesses failure in his attempt to "demarcate organisms from non-organisms within the class of functionally integrated entities."[155]

Wilson argues that asking whether something is an organism or not (e.g. an ant colony) is not to call for a conceptual analysis of the term "organism," nor suggest that an exploration of homology will provide an answer. The real question is whether "enough homoplastic commonalities can be found between these entities and the ones that we are all comfortable calling organisms."[156]

> The evidence does not suggest that (sic) the existence of a robust natural kind tied to other important properties at the level of the functional individual level. Any multicellular entity has to face a common range of problems, but the range of possible solutions is large enough that we should not expect to be able to generalize across kingdoms about the mechanisms of development.[157]

Queller and Strassman suggest that what we call superorganisms are either organisms, not organisms (just e.g. colonies) or perhaps, "quasi-organisms:"

> If superorganism is meant to highlight the common features of high cooperation, low conflict and unanimity of action, why reserve that term for the top level only? If instead the intent is to imply that these are somehow different, that they have not quite reached the level of organisms, we really ought to call them quasi-organisms.[158]

Like others, Matthew Haber is worried about the concept of an "organism," but considers the problem in regard to purported *super*organisms, namely eusocial insect colonies. Haber believes the term "organism" is unhelpful, and worse, muddies the waters. He is happy to

[155] Wilson, "Ontological butchery," 302.
[156] Wilson, "Ontological butchery," 309.
[157] Wilson, "Ontological butchery," 309.
[158] Queller and Strassman, "Beyond Society," 3148.

count colonies as individuals, but whether they are super*organisms* depends on what we mean by "organism." [159]

At this point in the chapter, we have many candidate terms (specific and general) to describe functionally integrated entities comprised of more than one evolutionary individual (i.e., organism, metabolic individual, immunological individual, holobiont, physiological individual). "Some individuals do not generate lineages, *per se*, but instead persist by way of containing lineage-generating individuals."[160] But if colonies can be evolutionary individuals that generate lineages of colonies, why not call them individuals (rather than the ambiguous "organism" or "superorganism")?

Haber gives an overarching label to the common problems with talk about individuals and organisms, "the problem of the paradigm," which is when we take an organism to be a paradigmatic individual, or when we use one kind of organism in particular as a paradigm organism "against which all others must be judged."[161] In other words, no one agrees on what an organism is exactly, nor which kind of the many different types of organism that colonies are meant to be similar enough to in order to be labelled a superorganism. "There is no better reason to define colony-individuals in terms of organisms, than to define organisms in terms of similarity to colonies."[162] Also, why consider colonies as similar or analogous to organisms, as some do, rather than a *kind* of organism?

The focus, Haber concludes, should not be on whether or how closely colonies resemble organisms, but rather on being units of selection and other biological as well as ecological aspects.

Karen Kovaka advances a critical response to Haber's "Problem of the Paradigm." Whereas Haber sees a solution in finding the right conception of biological individuality, Kovaka is dubious. Haber's concern is that we think of colonies as individuals to the degree that they resemble organisms. In other words, what "makes organisms individuals is necessarily, rather

[159] Haber, "Colonies are individuals."
[160] Haber, "Colonies are individuals," 212.
[161] Alternative wording: The problem of the paradigm is the "presumption that organisms are paradigmatic individuals, or the presumption of a paradigmatic organism against which all others must be judged" (196).
[162] Haber, "Colonies are individuals," 201.

than incidentally, due to their organismal characteristics."[163] But Kovaka replies that settling the individuality debate does nothing to help biologists who must still make use of analogical reasoning in a way that avoids the problem of the paradigm. Kovaka suggests that biologists can use similarities (e.g. between colonies and organisms) to come up with testable hypotheses, as opposed to working from, and making further, assumptions (as Haber fears).[164]

3.5.3 Extended Individuality

Influenced by the enactivist approach to living systems, Scott Turner offers a cognitive conception of individuality.[165] With a permissive notion of "cognition" articulated with thermodynamic descriptions of biological systems, Turner argues that an individual is one that has a "sense of itself as a system that stands apart in some unique way from the world around it." On his conception, even cells exhibit a minimal form of cognition, as it "must form an impression of the outside world and interpret it in a way that sustains the pool of low entropy that is the membrane-bounded cell."[166]

Turner refers to symbiotic organisms as "metaphorical organisms," "the physical manifestation of a metabolically convenient association between two complementary genomes."[167] Like O'Malley, Turner offers metabolic associations as explanations for the cohesion of symbionts rather than genetic relatedness. For example, termites are eusocial insects, yet are not characterized by the genetic relatedness of say, worker honeybees that share 75% of their genes (haplodiploidy). Instead of a genetic explanation (inclusive fitness), we can point to the exchange of intestinal fungi and bacteria essential for the digestion of wood, as an "impetus for sociality… enforced through the need to continually re-inoculate the group's myriad digestive tracts through the frequent exchange of oral and anal liquids."[168] This explanation appeals to metabolic dependence rather than genetic relatedness.

---

[163] Kovaka, "Biological individuality and scientific practice," 1100.
[164] E.g., Brian Johnson and Timothy Linksvayer, "Deconstructing the superorganism: social physiology, groundplans, and sociogenomics," *The Quarterly Review of Biology* 85, no. 1 (2010): 57-79.
[165] Enactivism holds that cognition is the product of one's interaction with the environment and is related to the notion of embodied cognition. Scott Turner, "Individuality in a Social Insect Assemblage," in Bouchard and Huneman, *From groups to individuals*, 219-241. See also William Morton Wheeler, "The Ant-Colony as an Organism," *Journal of Morphology* 22, no. 2 (1911): 307-325.
[166] Turner, "Individuality in A Social Insect Assemblage," 237.
[167] Turner, "Individuality in A Social Insect Assemblage," 237.
[168] Turner, "Individuality in A Social Insect Assemblage," 223.

Turner describes the termite mound as part of the colony's "extended physiology." The mound acts like a respiratory organ, circulating air and facilitating gas exchange like a giant lung. Termites respond to mound damage much like cells respond to tissue damage within a multicellular organism. The mound and colony exhibit homeostasis in response to external perturbation, and for Turner therefore count as an "extended organism."

I include Turner's account to represent purported individuals that may include intuitively non-living elements i.e. parts of the environment (e.g. earthen nests of ants or calcium carbonate of coral reefs).

3.5.4 Summary

In addition to having surveyed accounts of individuality which highlight or favor a certain sense or aspect of biological individuality, we have in this section now surveyed in more detail the kinds of cases in which various accounts may be of use. Extended individuals are cashed out in terms of metabolic and thermodynamic processes sometimes involving non-living components. Colonies are units of selection, and therefore evolutionary individuals regardless of their status as "organisms." Holobionts such as human beings and microbiota are usefully individuated by the immunological integration they share and comprise. We can roughly (but thoughtfully) relate these cases to the accounts featured above (in section 3.4).

We turn now to a pluralistic approach to biological individuality which offers various ways to hold a wider set of factors we want to include in our thinking about a range of biological individuals.

3.6 Pluralist Accounts

We have encountered different accounts which are an endorsement of some criteria over others, e.g. evolutionary or immunological considerations. But pluralist accounts want to recognize different types of individuals within the same theory. On this kind of framework, multiple types of individuality will coincide in some cases.

The main shift here is going from different accounts, conceptions, or senses of individuality, to a theory which accommodates what may otherwise be grouped into exclusive categories.

There is nothing new to doing so. Harper offered the terms “ramet” and “genet” to be of use when describing plant life.[169] The terms differentiate between a single clone and an entire clonal colony (the former a unit in the physiological sense, the latter a unit in a genetic sense, comprising all products of a single seed). Similarly, Hull recommended the terms “replicators” and “interactors” to make distinct two different types of biological entities we otherwise ambiguously refer to using the term “unit of selection:” the former picking out the units that are copied and transmitted (e.g. genes), and the latter picking out ecological units that are visible to selection such that their interaction with their environment leads to differential replication of replicators.[170]

Many of the foregoing philosophers mentioned in the previous sections of this chapter have a pluralistic view, that is, they allow for and articulate different types of individuals. Indeed, at least some version of pluralism seems to be the default, if implicit, view of most working in this area. For example, while O’Malley has a special interest in a particular type of biological individual (i.e. metabolic individuals), she pinpoints this subset by contrasting biological, Darwinian, and (what are traditionally called) organisms. Similarly, while what makes Pradeu’s account interesting is that it offers a picture of the individual as a primarily immunological entity, he makes a distinction between, e.g. physiological individuality and evolutionary individuality. In considering what follows, I should not be understood as saying only these among those I’ve featured in this chapter are pluralists. Rather I feature the following accounts here for their unique and novel concepts and terminology which in a more explicit way self-categorize under the banner of “pluralism.”

Godfrey-Smith asks us to think of “reproducers” at different levels in the biological hierarchy, providing a model for thinking about paradigm individuality in terms of degrees. Jack Wilson offers different kinds of individuals, each accommodating a set of our intuitions (or candidate criteria). DiFrisco takes a similar track to Wilson, offering his account as a preferable alternative to Clarke’s monist account.

3.6.1 Godfrey-Smith’s Account

[169] Harper, "Population biology of plants."
[170] David Hull, "Individuality and selection," *Annual review of ecology and systematics* 11, no. 1 (1980): 311-332.

As we have seen, the living world bears a mix of metabolic and reproductive properties (among others). Peter Godfrey-Smith relates the former to "organisms" and the latter to "Darwinian individuals."[171]

Those like Clarke, who offer definitions of evolutionary individuality, do so by drawing on the concept of a Darwinian population – a population that evolves in that its members undergo natural selection (i.e. because members exhibit heritable variation in fitness). Thus, an evolutionary individual can be defined as that which is a member of a Darwinian population. So long as biological entities are members of a population such that they are capable of undergoing natural selection, they are individuals in this Darwinian sense. Counter-intuitively then (as mentioned previously in section 3.3.1), these kinds of biological individuals may not even be living (e.g. genes, groups).

On the other hand, we might have organismic individuals that are not participants in natural selection, for example in the case of symbiotic, multi-lineage units ("collaborations" in Dupre and O'Malley's sense). Those biological entities often considered to be paradigm individuals are considered to be both organismal persisters and Darwinian individuals.

Note that with pluralist accounts the "what is important" component has many answers. Somewhat like Clarke's monist account, Godfrey-Smith develops a way of approaching the problem of biological individuality by considering not only different criteria (to various degrees) but different kinds of individuals. In this regard, however, Godfrey-Smith addresses the topic of individuality somewhat indirectly, by focusing on different types of "reproducers"[172] abstracted from three paradigm, differently "levelled," individuals: organisms, cells, and genes (collective reproducers, simple reproducers, and scaffolded reproducers, respectively). Individuals ("reproducers") are individuals in the sense that they, together with others of their kind, form Darwinian populations (the criteria for which is Godfrey-Smith's' main focus).

Collective reproducers are

[171] Peter Godfrey-Smith, "Individuality, subjectivity, and minimal cognition," *Biology & Philosophy* 31, no. 6 (2016): 775-796.
[172] A term also used by James Griesemer.

> reproducing entities with parts that themselves have the capacity to reproduce, where the parts do so largely through their own resources rather than through the coordinated activity of the whole. Not *all* the parts need to be able to do this for an entity to count as a collective, the requirement is that some can.[173]

Examples include "buffalo herds and other social groups, multicellular organisms like ourselves, symbiotic associations that are not too tight, and colonies."[174]

Simple reproducers are the "lowest level entities" that can produce "using their own machinery, in conjunction with external sources of energy and raw materials."[175] The entity's "machinery" is "not a collection of things that are able to reproduce under *their* own steam. Cell division is an activity of the whole cell."[176] Examples include bacterial cells.

Scaffolded reproducers are entities

> which get reproduced as part of the reproduction of some larger unit (a simple reproducer), or that are reproduced by some other entity. Their reproduction is dependent on an elaborate scaffolding of some kind that is external to them. However, these entities do have parent–offspring relationships, hence they form lineages or family trees.[177]

In contrast to simple reproducers, scaffolded reproducers do not reproduce without metabolic assistance; they lack the "machinery" required to reproduce. Examples include viruses and chromosomes.

[173] Godfrey-Smith, *Darwinian populations,* 87.
[174] Godfrey-Smith, *Darwinian populations,* 87.
[175] Not lowest on the hierarchy, but lowest that are self-sufficient in their capabilities to reproduce (i.e. a bit of viral DNA may be lower on the hierarchy but relies on a cell, in which it must infect and hijack, for reproduction).
[176] Godfrey-Smith, *Darwinian populations,* 88.
[177] Godfrey-Smith, *Darwinian populations,* 88.

To sum up, collective reproducers may have producers of all three kinds as constituent parts (scaffolded, simple, and collective). Simple reproducers may only have scaffolded reproducers as parts. Scaffolded reproducers cannot have reproducers as parts.[178]

Focusing on "collective reproducers" (and therefore, with "organisms" in mind), Godfrey-Smith lays out several criteria with which he builds a multi-dimensional model in which to plot different cases (along three sides of the cube seen in figure 3.1).[179] The idea is to represent candidate criteria as a continuum along some axis. As mentioned previously, some accounts of individuality make use of, or focus on, a single criterion.[180]

A bottleneck (B) is a "narrowing that marks the divide between generations." This admits of degrees. The tighter the bottleneck, the "less the new structure is a mere continuation of the old."[181] Examples of high degrees of B include a single germ cell, which marks the initial stage of that which will develop into a fully formed adult organism, or a runner from one ramet to another (e.g. strawberry plant).

The variable "G" stands for germ line - "the degree of reproductive specialization of parts, in the sense involved in germ/soma distinctions and related phenomena."[182] This admits of degrees. Examples of high degrees of G include the separation of mammalian sex cells and somatic cells; only the former can produce another organism. Another example is the division of reproductive labor in a honeybee colony, in which the queen and worker bees may be analogous to germ and somatic cells.

Lastly, Integration (I) is "the extent of division of labor, the mutual dependence (loss of autonomy) of parts, and the maintenance of a boundary between a collective and what is outside it."[183] Loose collectives have low levels of I; tightly bound collectives possess medium levels; complex, adapted organisms possess high levels. Examples of high degrees of I include an Oak ramet.

---

[178] Godfrey-Smith is careful to note there are cases that exist outside and in between these idealized categories: "The eukaryotic cell is a former collective, and one that still has some features of a collective. Mitochondria, in different organisms, are at various locations on a road between simple and scaffolded."
[179] In his book, Godfrey-Smith features nine variables – the model permits three at a time.
[180] E.g. Buss emphasizes germline sequestration, Dawkins emphasizes evolutionary bottlenecks, and Wilson emphasizes functional integration.
[181] Godfrey-Smith, *Darwinian populations,* 92.
[182] Godfrey-Smith, *Darwinian populations,* 92.
[183] Godfrey-Smith, *Darwinian populations,* 93.

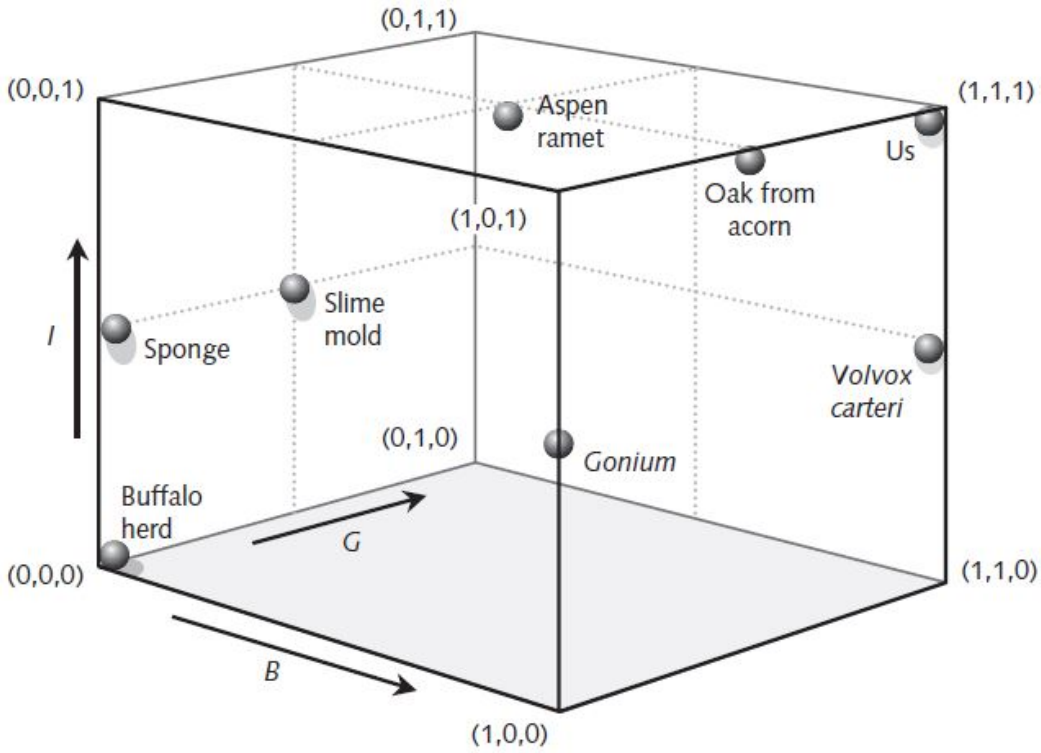

Figure 3.1: Godfrey-Smith's multi-dimensional model of collective reproducers.[184]

These collective reproducers are individuals to the degree that they exhibit B, G, and I.[185] Some cannot be profitably thought of as individuals. Others will be paradigm individuals. Others still may be marginal cases.

So, for Godfrey-Smith we have not only a plurality of individuals ("reproducers"), but a spectrum of more or less realized paradigm individuality.

Godfrey-Smith's focus on reproducers and lineages can accommodate and highlight some of the biological realities emphasized by Pradeu and O'Malley. Many of the paradigm individuals that constitute lineages are perhaps not the clear "reproducers" we take them to be. As the literature on holobiionts should make clear, "collective reproducers" like human beings are "a local nexus of different lineages of reproducing entities."[186] This is a feature and not a bug for Godfrey-Smith's model, which is not only useful for sketching a pluralist

[184] Reproduced from Godfrey-Smith, *Darwinian populations,* 95.
[185] We could model other candidate criteria as well (three at a time).
[186] Pradeu, "Organisms or biological individuals?"

approach to individuality – it also allows us to represent various criteria met *to a degree* (i.e. between 0 and 1).

In conclusion, Godfrey-Smith's account shows how different criteria which sometimes feature in conflicting monist accounts, can be combined to give an overall rating or score of paradigmatic individuality. This means that like Clarke's monist account, even pluralist accounts can articulate the notion of individuality coming in degrees.

In section 5.1.5, I will suggest a useful way to capitalize on some of Godfrey-Smith's conceptual work by applying his multi-dimensional models to candidate agents (including collectives), and criteria for agency.

3.6.2 Wilson's Pluralist Account

Jack Wilson also offers a pluralist account. Rather than Godfrey-Smith's "reproducers" which roughly map onto "levels" of the evolutionary hierarchy, Wilson gives us a picture of overlapping individuals, i.e. different individuals picked out according to different delineated criteria: genetic individuals, developmental individuals, and functional individuals.

Wilson notes six things, one or more of which we take to be true when calling a living thing an "individual:"

1. It is a particular. A biological entity is a particular just in case it is neither a universal nor a class.
2. It is a historical entity. A biological entity is a historical individual if it is composed of spatiotemporally continuous parts.
3. It is a functional individual. A biological entity is a functional individual if the parts which compose it are causally integrated into a functional unit.[187]
4. It is a genetic individual. A biological entity is a genetic individual if all of its parts share a common genotype.
5. It is a developmental individual. A biological entity is a developmental individual if it is

[187] Wilson notes limits of this kind in regards to picking out "organisms": "I have tried to demarcate organisms from non-organisms within the class of functionally integrated entities but have not found a satisfactory answer." Wilson, "Ontological Butchery," S302.

the product of a developmental process.
6. It is a unit of evolution. A biological entity is a unit of evolution if it functions as an important unit in an evolutionary process.

Paradigm individuals, e.g. "higher animals," will be all six.

Wilson's pluralist account then, gives us categories in which we can be more specific about what we mean by "individual" in different contexts. Specifically, 3-6 serve to help achieve a more nuanced, less ambiguous approach to biological individuality. While the terms are not new, Wilson distinguishes the senses of each for his account.

To see how these come apart, imagine a hypothetical situation in the context of a future in which living forms created in the lab by way of "bottom-up" synthetic biology (especially a futuristic science-fiction version, e.g., molecular assemblage technology).

Genetic individuals have parts that "share a common genotype because of descent without interruption from a common ancestor with that genotype."[188] The latter bit precludes certain cases in which we could create a genetic individual from scratch. On a genetic understanding, such entities could not be individuals.

A functional individual operates as a "functional organic whole" composed of "causally integrated heterogeneous parts." Functional individuals are not classified based on their history or genetic composition, only the at-present causal relation of their parts.[189] Biological entities built in our hypothetical lab could be individuals in this sense.

Developmental individuals begin from a bottle neck (a single or few cells) and "will, unless interrupted, develop into a multicellular body."[190] Biological entities created in our lab may or may not be individuals. If what was created from scratch was a single (or few) cells that could, and was then allowed to, multiply and develop, it would be an individual. However, if an adult creature was assembled atom by atom (and presumably "zapped" to life somehow –

[188] Wilson, *Biological individuality,* 86.
[189] Wilson, *Biological individuality,* 89.
[190] Wilson, *Biological individuality,* 99.

but let's not let that distract us), it would not be an individual in this sense.[191]

Thus, in the spirit of Harper's "genet" and "ramet" and Hull's "replicators" and interactors," Jack Wilson suggests a plurality of individual terms to map onto the delineated biological entities that concern us for various purposes. The upshot is that on this view, we have different (though sometimes overlapping) individuals we can use with more clarity and specificity.

3.6.3 DiFrisco's Pluralist Account

Like Jack Wilson, DiFrisco offers a trio of biological individuals: ecological, physiological, and developmental. But to understand DiFrisco's unique offering, we must understand the motivation for it. I give special attention to DiFrisco's account as it sheds light on some of the problems with other accounts, particularly Clarke's evolutionary account.

Many philosophers stress that different criteria yield different answers to the question, *is x an individual?* For example, a biological entity may be a unit of selection yet lack its own metabolism; it may be genetically homogeneous but lack physical contiguity; it may exhibit adaptations but lack a bottle-necked developmental origin. Clarke's monist account unifies these various criteria, proposing that they can function as types of individuating mechanisms. That is, what makes an individual can be realized in a number of ways (i.e. the functions that constitute biological individuality have a range of different types of instantiation).

The alternative of course, is to recognize a plurality of individuals. DiFrisco takes this approach, noting that by "pluralism" we should mean *kind pluralism*, not just *criteria pluralism*.[192]

In considering various criteria of individuation, DiFrisco argues that we should be aiming to match them up to our theoretical goals. This "mapping" of criteria to *desiderata* is what an account of biological individuality *is* (or should be). DiFrisco lists nine commonly cited aims:

[191] These simplified examples are no substitute for Wilson's careful treatment of each category. The interested reader is referred to chapter 5 of his book.
[192] DiFrisco, "Kinds of biological individuals," 870.

1. Measure reproductive fitness (i.e. count offspring)
2. Measure population size and track demographic change
3. Distinguish reproduction from growth
4. Guide our choice of selection models (e.g. choosing focal "levels")
5. Identify evolutionary transitions in individuality
6. Predict histocompatibility, fusion, and transplantation outcomes
7. Provide biological kinds for classification and inductive generalization (i.e. what it is we are comparing across taxa)
8. Determine the appropriate reference system for decompositions into functions and characters
9. Determine the unit of ecological interaction (individuals that participate in ecological processes as a whole, i.e. demarcate "interactors")[193]

For some this list may be longer or shorter. The longer it grows, DiFrisco argues, the less plausible a monist account becomes (i.e. the less likely a single criterion will "map onto" the entire list). For example, an account like Clarke's (an evolutionary account based on functional criteria) may clearly satisfy aims 1-3, but not aims 7-9. DiFrisco's pluralist solution is to offer three sub-kinds of biological individuals: physiological individual, ecological individual, and developmental individual.

DiFrisco's pluralism is "kind pluralism," in which "there are biological individuals wherever there are countable individuals instantiating biological properties that are projectable in a sense relevant to aims (1)– (9)."[194] The phrase "biological individual" refers to a "class containing heterogeneous individuative and explanatory kinds." DiFrisco wants to emphasize to the reader that "biological individuality is as much about kinds as it is about individuals."[195]

DiFrisco is critical of functionalist evolutionary accounts and of monist accounts more generally. He therefore makes his case by taking aim at Ellen Clarke's account.

> Functional kinds like 'evolutionary individual' need to be mapped onto identifiable material traits… in order to enable counting their instances. Since these material traits

[193] This is an edited version of DiFrisco's list.
[194] DiFrisco, "Kinds of biological individuals," 853.
[195] DiFrisco, "Kinds of biological individuals," 849.

> can malfunction, however, the entities possessing them might fail to be instances of the functional kind. Hence, possession of these traits cannot constitute a sufficient condition for the functional capacity in the way required by the functionalist view of evolutionary individuality.[196]

Whereas Clarke hopes to improve upon and unify accounts that suggest specific traits ("material realizers") by suggesting functional criteria, DiFrisco alternatively suggests a disjunction of those traits.[197] In other words, instead of capturing these various traits under the functional catch-all "individuating mechanisms," we might instead offer an account of evolutionary individuality in which individuality is satisfied by, e.g. bottle necked life cycle *or* genetic homogeneity *or* an immune response, etc. This solution scratches the itch that motivated Clarke's multiply realizable account, but is not open to the same criticism.

Yet uncertainty remains. To address the previously cited problem (the possibility of individuating mechanisms malfunctioning), DiFrisco suggests *stipulating* those material realizers (the presence of those traits) as individuating mechanisms, and gauge degrees of individuality based on probabilities (i.e. never sure if the individuating mechanism is functioning properly, but reasonably grounded statistically, in that 1 – x cases will exhibit proper functioning). It's not perfect, but at least it is operational.[198] (However, DiFrisco will conclude that even the proposed disjunctive version of the functionalist account cannot satisfy theoretical *desiderata* 7-9).

DiFrisco is keen to point out the insufficiency of an evolutionary account (even one amended as he proposed). Consider an evolutionary account's inability to deal with the fundamental question raised by aim 7: what are we comparing across taxa? He invites us to compare a European honey bee colony with an individual carpenter bee. Both are evolutionary individuals (the former *colony* and the latter *bee*).[199] This is counter-intuitive as we observe obvious homologous traits when comparing bee to bee. When comparing colony to bee on an

---

[196] DiFrisco, "Kinds of biological individuals," 858.
[197] That is, *inclusively* disjunctive.
[198] Without this probability-based approach, we could not be sure if the criteria were met without first knowing if the criteria were met (so not very useful).
[199] For how a colony can be a unit of selection see e.g. Buss, *The evolution of individuality*; Thomas Seeley, *The wisdom of the hive: the social physiology of honey bee colonies* (Harvard University Press, 2009); Francis Ratnieks and P. Kirk Visscher, "Worker policing in the honeybee," *Nature* 342, no. 6251 (1989): 796-797.

evolutionary approach, "the shared trait functions are largely functions relating to reproduction."[200] Are there alternative frameworks to make sense of these important similarities (e.g. germ-soma separation)? DiFrisco suggests that rather than ignore our intuitions about the qualities that better help us satisfy aims e.g., 7-8, we could instead reappraise the colony-level individuating mechanisms. For example, germ-soma separation at the colony level can perhaps more appropriately be seen as a dominance hierarchy observed in other social animals (i.e. queen asserting reproductive dominance over workers).[201]

Recall that Clarke justifies her view by citing the great utility of an evolutionary account's power to predict and explain evolutionary processes. Yet, as demonstrated with the comparison of bee colonies and bees, comparing evolutionary individuals may "not be as sound a basis for informative comparisons as more structurally based factors."[202] To claim that an evolutionary approach holds the most explanatory value "assumes that the force of selection overcomes differences due to structure and level of organization to produce convergent similarities—an adaptationism about individuality."[203]

Arguably, the upshot supports the case for pluralism. Suppose Clarke is right about what motivates her position, that an evolutionary account is the best on offer when choosing *one* account over others - it does not follow that we should adopt *solely* an evolutionary approach.

This is what motivates DiFrisco's offering of three non-evolutionary kinds, which are developmental, physiological, and ecological individuals. These terms mostly explain themselves, but I will provide some brief exposition.

Physiological individuality conforms to our intuitions about phenomenal chunks of the living world. Mason bees and honey bees can both be this kind of individual. More specifically, physiological individuality can account for sterile workers of eusocial colonies (e.g. bees, wasps, naked mole rats). They will exhibit low degrees of evolutionary individuality (think of

[200] DiFrisco, "Kinds of biological individuals," 863.
[201] DiFrisco, "Kinds of biological individuals," 862.
[202] DiFrisco, "Kinds of biological individuals," 863.
[203] DiFrisco, "Kinds of biological individuals," 863.

Godfrey-Smith's three-dimensional model), yet still are "subject to metabolic scaling laws," and exhibiting "organizational closure"[204] just as their non-sterile kin.

Ecological individuality can account for holobionts (e.g. host organism and microbes, lichen, etc.) and extended phenotypes (e.g. spider's web, the "reef" of corals). These heterogeneous units are often the focus of individual-based modeling in ecology.

Developmental individuals are "a subset (or superset) of evolutionary, physiological, or ecological individuals that is produced by a series of cell divisions contiguous with a bottleneck, budding, or fragmentation event."[205] Developmental individuality is "the most relevant criteria for assessing factors like whether an entity has a lifespan and undergoes developmental processes such as maturation and senescence."[206] This type of developmental individuality helps meet our aims for projectability based on homologies, as "due to the conservatism of development… developmental individuals often most reliably constitute the bearers of homologous traits."

Whereas an account like Clarke's may be most appropriate for aims 1-3, DiFrisco offers these three types of individuals as appropriate for aims 7-9. It is DiFrisco's position that "'biological individual" and "organism" are classes containing individuative and explanatory kinds, but are not themselves individuative and explanatory kinds."[207] DiFrisco makes a strong case for pluralism, arguing that "the problem of biological individuality is posed at too high a level of generality, and that theoretical unity can be re-discovered at a finer grain."[208]

3.7 Taking Stock

In this chapter, we have surveyed a few accounts, noted a number of candidate individuals, and observed the range of individuality criteria on offer. Many will feel that the most sensible way forward is to adopt a pluralistic view about biological individuality. (Indeed, we have noted the recent trend in the literature to speak broadly of evolutionary individuals and

[204] See Álvaro Moreno and Matteo Mossio, *Biological Autonomy: A Philosophical and Theoretical Enquiry* (Dordrecht: Springer, 2015).
[205] DiFrisco, "Kinds of biological individuals," 867.
[206] DiFrisco, "Kinds of biological individuals," 868.
[207] DiFrisco, "Kinds of biological individuals," 870.
[208] DiFrisco, "Kinds of biological individuals," 871.

physiological *individuals*, as opposed to evolutionary or physiological *accounts* of individuality). There are a number of ways to delineate biological individuals according to our purposes. These individuals will overlap unevenly. Of course, this overlap will be at least somewhat unsatisfying. Why? because the total number of "biological individuals" will not map to our primary *desiderata*: chunks of the living world meaningfully individuated across a range of purposes in line with our intuitions.

The biological debate largely concerns itself with curious cases. Should the Portuguese man 'o war, a clonal strawberry plant, a termite colony, and a slime mold be considered individuals? The problem of in-between cases and multiple criteria can be made distinct, and yet some forms of pluralism (e.g. Godfrey-Smith's) shows us how they are intertwined. If best *descriptions and explanations* of these cases call for an integration of what we can make logically distinct in our analysis, it might not be useful for those explanations to rely on individuality concepts! Consider also that some curious cases are not general, but particular cases. In the following chapters I will illustrate why agential explanations are useful in such cases (e.g. male plant sterility or cancer in some multicellular organisms). In 5.2, I will argue that enumerating a number of different kinds of individuals does not capture what is salient about curious cases. Nor, I will suggest, does allowing that individuals can be individuals to a degree (whether on a monist or pluralist account). So perhaps then, even pluralism is not doing all the work we want it to do. For some of our theorizing, pluralism about biological individuality may prove unsatisfying for the very reasons it is offered. I will argue in chapter 5 that the concept of agency is useful in providing fuller explanations for the kinds of cases on which the debate has sometimes focused. I will suggest that for *our explanations* of the salient phenomena those in the debate have taken interest in, "as-if" agency heuristics are readily available to us, even while the jury is out on what counts as an individual (of some kind at some time). Counter-intuitively then, if we want to know what is curious about "weird" cases of purported biological individuality, we will not get there by adding to or refining our existing concepts of individuality alone. We should also keep the concept of agency front and center. I will develop these points further in chapter 5.

Looking in the other direction, my explicit purpose in this thesis is not to attempt to settle either debate from within (biological individuality or group agency). Rather I want to show how the problems of one debate could assist in propelling the other debate towards a solution to a particular problem it faces. So, as I will suggest in section 5.1, regardless of what we

think about pluralism in the context of the present debate (biological individuality), the pluralistic perspective about individuality might be usefully applied to our social categories in moral and political philosophy to best preserve important differences that matter to us.

Chapter 4
The Agential Perspective – Building the Bridge between Group Agency and Biological Individuality

In this chapter I aim to build a bridge between the two areas of philosophy we have so far explored in this thesis: philosophy of biology and moral and political philosophy. Specifically, I will argue that the concept of agency, as an achievement of unity, may serve as a legitimate and potentially fruitful common currency between discussions of biological individuality and group agency. If successful, what is presented in this chapter will justify importing and exporting some concepts to and from both bodies of thought. In the following two chapters, I will offer some suggestions as to how mutually profitable commerce may commence by way of the bridge established in this chapter.

In Section 1, I will explain why teleological language in biology is seen on the one hand as taboo, and on the other as indispensable. This will motivate a need to distinguish what is useful from what is misleading about teleological language. In Section 2, I will explore the various meanings of "teleology." I will demonstrate how we can naturalize concepts otherwise considered problematic when used in the context of biology, and further, that there is utility in deploying these concepts. I will conclude that some uses of teleological language in biology are perfectly acceptable, if not indeed indispensable. In Section 3, I will introduce the heuristic of "agential thinking" as a form of adaptationist reasoning – a way of conceptualizing biological individuals as agents pursuing the goal of fitness maximization. In Section 4, I will argue that both literally intended, and heuristically intended ascriptions of agency are justified by the unity exhibited by the targets of those ascriptions. In Section 5, I will suggest some methodological commonalities, in both the features and the problems, between the philosophy of biology and moral and political philosophy. In Section 6, I will offer a summary, concluding that we may proceed with confidence, free to traverse the established bridge between sub-fields.

4.1 The Taboo on Teleological Language in the Natural Sciences

In this section I will discuss why teleological language is thought to be out of place in the natural sciences, but why omitting it completely from biology is also problematic.

A first-year university student in biology will learn very quickly to be careful when thinking and speaking about the natural world in a way that might otherwise be unproblematic. We often speak of things happening *for reasons or purposes*. This form of explanation often assumes or implies a goal or purpose for some observed phenomenon. This is quite natural when the phenomenon in question is an action (e.g. *I started the car in order to drive it to the store in order to buy some bread*). However, when our explanations of the natural world take this form, what may be implied (intentionally or unintentionally) or assumed (as it was historically) is a divine actor, purpose, or force directing things towards a pre-ordained end state. In addition to the problem of implying an actor where there is none is the problem of implying human-like agency where there is none (e.g. consciously or not, conceiving of a bacterium as possessing a sense of purpose and conscious goals).

The scientific revolution is often marked by the rise of a mechanistic view about the physical world. More recently, Darwin's work can be seen as having extended this perspective to the living world. But vitalistic intuitions remained well into the 20th century, haunting biologists and prolonging the exorcism of teleology's ghost from scientific thinking.

It is therefore seen as a delicate subject to point out that, although the *literal* application of teleological concepts to nature has been rejected, teleological language often seems unavoidable when trying to give adequate and useful explanations about certain biological phenomena (and I will provide an example below).[1] In view of this fact, it would be tantamount to superstition to disallow ourselves a useful and powerful, if not indispensable, tool for our engagement with the living world. Further (and highly relevant to material in the following chapter), such tools can be usefully applied at various levels in the biological hierarchy, as pointed out by Ernst Mayr: "the heuristic value of the teleological makes it a powerful tool in biological analysis, from the study of the structural configuration of macromolecules up to the study of cooperative behavior in social systems."[2]

So then, despite giving rise to much confusion, teleological language plays an important role in biological explanations. We should, I believe, back up and make the case that judicious use

[1] E.g. Francisco Ayala, "Teleological explanations versus teleology," *History and philosophy of the life sciences* 20, no. 1 (1998): 41-50.
[2] Ernst Mayr, "Teleological and Teleonomic, a New Analysis," *Boston Studies in the Philosophy of Science* 14 (1974): 157.

of teleological language is justified. We should make the case both that 1). Teleological language can be suitably naturalized and demonstrably deployed in an unproblematic way, and 2). There is added utility to deploying teleological concepts when speaking and reasoning about natural processes.

Before giving an overview of the case for (1), I believe we can quickly make the case for (2) by way of example. Here I will use an example similar to one given by Mayr.[3] Consider this articulation of the adaptedness of an animal's behavior: *The bird migrates in autumn in order to escape the bad weather and food shortages of the north.* If not read correctly as adaptationist shorthand, the phrase "in order to" may be misinterpreted as conveying foresight and the intentional pursuit of some end (i.e. what a first-year biology student has been taught to carefully avoid). This problem is made more obvious in cases that feature traits such as evolved anatomical structures, e.g., *the moth's color pattern is for camouflage*. Mayr clarifies:

> Darwin has taught us that seemingly teleological evolutionary changes and the production of adapted features are simply the result of variational evolution… and the probabilistic survival of those individuals with the temporarily fittest phenotype… Adaptedness thus is an *a posteriori* result rather than an *a priori* goal seeking.[4]

However, when articulated in a manner devoid of teleological language - *the bird migrates in the autumn and thereby escapes the bad weather and food shortages of the north* - we seem bereft of an answer to the question implicit in the original explanation: *why does the bird migrate?*

Examples like this demonstrate how a strict avoidance of teleological language robs us of a tractable form of explanation that effortlessly homes in on the relevant *explananda.* But the example also leaves us with questions about the relationship between explanations and causality. *Why does the tortoise have a shell? Why does the bird preen its feathers?* Are we soliciting an answer in the form of a causal explanation? In what follows, I will answer "yes," but with nuance and detail that "yes" alone could not provide. So, let us move on to whether

[3] Mayr, "Teleological and Teleonomic."
[4] Ernst Mayr, "The idea of teleology," *Journal of the History of Ideas* 53, no. 1 (1992): 131.

teleology can be naturalized. To do so we must attend to "teleology" itself, and in doing so, unpack several common senses or usages. Then we can be clear about what kind of explanations we are aiming to capture with our biological analyses articulated in teleological terms.

4.2 Purposes, Goals, and Functions – Actual or "as if"

In this section I will explore the various meanings of "teleology" for the purpose of distinguishing what is useful from what is misleading about using teleological language in biology. This discussion will also lead us into an introduction of the notion of "biological functions."

4.2.1 The Meanings of "Teleological"

Telos refers to an end or goal, and since Aristotle, the notion of "final causality" is often expressed with the related notion of purpose.[5] We can immediately see the problem of using teleological language when providing causal explanations. Familiar examples of causality often illustrate what you might call "spatial pushing" - one billiard ball knocks into a second ball propelling the second ball forward. Thus, we conclude the first billiard ball caused the second to move. But the notion of *final causality* has often been thought to be more like "temporal pulling." An example of what I'm calling "temporal pulling" might be something you would hear from a child: *Rain falls to make the flowers grow,* as if the flowers need for water played a causal role in rainfall. Conceptualizing final causation as "temporal pulling" risks bringing with it what is sometimes called "backward causation" – the notion that something in the future could cause something in the present (i.e. future states or events determining previous states or events such that future states are fulfilled according to some purposive, providential plan that previous states enact – hence "pulling" the present towards the future goal state). For obvious reasons, *naturalists* do not take seriously that the future can cause the present, nor the idea that rain exists as part of a grand plan to sprout future flowers.

[5] Cf. Marjorie Grene, "Aristotle and Modern Biology," in *Topics in the Philosophy of Biology*, eds., Marjorie Grene and Everett Mendelson (Dordrecht: Reidel,1976).

We have pinpointed two ideas associated with "teleology" that biologists would like to distance themselves from: backward causation and purposive directedness. When giving a causal explanation, e.g. for why a bird migrates or why humans have gallbladders, we do not want to invoke either of those ideas. The future does not cause the present and there is no reason to believe in an externally imposed, providential force guiding events towards some final fulfillment of purpose. This means evolution itself is not a process suitable for description in teleological terms. Darwin showed that when certain conditions were in place, evolution (by natural selection) was inexorable. There is therefore no need to posit some intrinsic force driving evolution towards a state of perfection (i.e. orthogenetic evolution). The evolutionary changes we observe are not "pulled" into the future nor "directed" by some force or deity. So, when we ask our question – *why do deer have antlers?* – any answers that cite as a cause either directedness or backwards causation are not what we are looking for.[6] But we need not conclude that all teleological language is inappropriate for use in biological explanations.

Biologists and others may find it useful to think and speak teleologically without making assumptions about nature, as a whole or any of its parts or aspects, progressing with an intrinsic directionality toward some ultimate goal.[7] Sometimes final causes are invoked to point to something's "purpose" without any commitment to backward causation. Sometimes the notion of a final cause is invoked *only* to provide an explanation, give a reason, or predict similar future instances of the phenomenon in question - used as a rule of thumb for making predictions without regard to purpose or directedness whatsoever.

This latter notion is a familiar one in biology - an "as if" kind of teleological reasoning. That is, we can look at the natural world and describe things *as if* they were fulfilling a purpose. With living organisms, for example, it might be helpful in making sense of organisms' behaviors (and predicting future behaviors) if we think about them as agents acting to reach a goal or fulfill a purpose. This is not limited to behaviors, but can even apply to anatomical features (e.g. a plant's thorn, a fish's fin), for which we can reason about as if physical traits

[6] Again, here I mean evolution as a process being directed towards some ultimate end ("meant to" bring about antler in deer). Goal-directed behavior of complex organisms may be "directed" in another, unproblematic sense.
[7] E.g. "The Scale of Perfection" or "Great Chain of Being" as expressed by NeoPlatonists and their forebears. Arthur Lovejoy, *The Great Chain of Being a Study of the History of an Idea: The William James Lectures Delivered at Harvard University, 1933* (Cambridge: Harvard University Press, 1964).

were strategies to fulfill an end (e.g. to survive and maximize fitness - an idea which will play a central role in sections 4.2 and 4.3). We might apply this as-if heuristic to a range of very different critters in the world: bacteria, worms, starfish, lizards, mushrooms, wolves, or viruses.

In an attempt to distinguish these different notions, some have suggested the alternative term "teleonomy" to denote an "as if" notion, i.e. *as if* the processes referred to were teleological. Unfortunately though, the term "teleonomy" has itself been variously used throughout the last century.[8] Hayne Reese, [9] working in Behavior Analysis, notes that the term "teleonomy" was introduced "seemingly independently" with non-identical usage by three different scholars: Allport,[10] Pittendrigh,[11] and Monod.[12] The problem of course, is that even "as-if" descriptions of various processes as "teleonomical" may be motivated by very different cases that are misleadingly lumped together. (Lumping things together is not helpful - we have already observed that the notion of final causality may or may not imply backward causation). So even though we have available a pre-existing proposed alternative term ("teleonomy"), its varied usage brings with it near the amount of baggage as "teleology."

We had better be clear on the various cases we would describe in teleological terms, and what further nuance they demand of us. I will therefore proceed with a treatment (drawing on Mayr's own) of the senses and usages of teleology (in which "teleonomy" will feature as only one, used more specifically). When all is said and done, I believe the *actual versus as-if* distinction will prove sufficiently useful when applied with more specificity to cases in which we heuristically invoke intentional attributions. Working through the various understandings of teleology is a route to that end.

---

[8] Mayr, "the idea of teleology," 127.
[9] Hayne Reese, "Teleology and teleonomy in behavior analysis," *The Behavior Analyst* 17, no. 1 (1994): 75-91.
[10] Floyd Allport, "Teleonomic description in the study of personality," *Character & Personality; A Quarterly for Psychodiagnostic & Allied Studies* (1937), 202-214.
[11] Colin Pittendrigh, "Adaptation, natural selection, and behavior," in *Behavior and evolution, eds.,* A. Roe and G. G. Simpson. *Behavior and evolution* (New Haven: Yale University Press, 1958).
[12] Allport in particular stresses the "as if" sense previously alluded to. Reece writes, "Pittendrigh said that teleonomy is functional when it refers to goal-directed behavior and evolutionary when it refers to the origin of this function, and in neither sense is it causal in the sense of an antecedent, productive cause." (83). So then, like Allport, Pittendrigh's teleonomy is about explaining and understanding phenomena, not positing some future cause of a process in the past. Monod's teleonomy is like Mayr's "program," or as Reece says "a blueprint for the future." Monod's notion of "program" is mysterious, perhaps mystical, but not finalistic in the sense indicating backward causation. Jacques Monod, *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology* (London: Collins, 1972).

Mayr distinguishes five classes of phenomena which have been described as teleological: Cosmic Teleology, Teleomatic processes, Teleonomic processes, Purposeful Behavior, and Adapted Features.[13]

We have already encountered the first kind of teleology: "cosmic teleology" (or "Grand Design Teleology") [14] – the idea that the natural world is progressing toward some ultimate purpose or perfect state (characteristic of some theological and Hegelian views). Teleological language used to invoke these ideas has no place in evolutionary biology.

The second class of phenomena said to be teleological are "teleomatic" processes. Charles Pierce suggested the term "finious" for natural inorganic processes' "tendency towards a final state" - processes that "tend asymptotically toward bringing about an ultimate state of a thing."[15] Peirce's concern is that we don't spuriously apply the term "teleological" to anything that reaches an end state as determined by properties in an initial state.[16] For example, according to the second law of thermodynamics, we would have to say that the entire universe is teleological in that entropy increases and culminates in equilibrium. This is particularly ironic considering that the heat death of the universe is a far cry from the ideas of those who would use the term "teleological" to express that the universe is intrinsically driven on a path of inevitable progress towards some ultimate and supreme purpose, perfection, or unity.[17] What Pierce calls finious, Mayr calls "teleomatic." Cups of coffee that cool towards room temperature and rivers that empty into tributaries can be described as teleomatic and avoid the associations with purpose or planning (i.e. end-directed in some sense, but not goal-directed). Teleomatic processes are those following natural laws – they are not biological phenomena. Teleomatic processes are not those we are concerned with when asking for explanations of biological phenomena for which teleological language might offer a concise answer (e.g. *why do ants have mandibles?).*

---

[13] Ernst Mayr, "The multiple meanings of teleological," in Towards a new philosophy of biology: Observations of an evolutionist, ed., Ernst Mayr (Harvard University, Press, 1988), 38-66.
[14] Reese, "Teleology and Teleonomy in Behavior Analysis."
[15] Charles Hartshorne, *Collected Papers of Charles Sanders Peirce*, eds., Paul Weiss and Arthur W. Burks (Harvard University Press, 1958); T.L. Short, "Peirce's Concept of Final Causation," *Transactions of the Charles S. Peirce Society* 17, no. 4 (1981): 369-82.
[16] C. H. Waddington, *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology* (London: Allen & Unwin, 1957).
[17] There may be room for a different view, e.g. if we subscribe to alternate forms of "radial" energy ala Pierre Teilhard de Chardin, or in contrast to the heat death hypothesis, the "big crunch" eschatology turns out to be right. See John Barrow and Frank Tipler, *The Anthropic Cosmological Principle* (New York: Oxford University Press, 1986).

Teleo*nomi*c processes, a third category, by contrast, are governed by materially-instantiated algorithms.[18] For Mayr then, "teleonomic" should only be substituted for "teleological" in instances when the latter is describing a process that "owes its goal-directedness to the operations of a program," with program defined as "coded or prearranged information that controls a process (or behavior) leading it toward a goal."[19] A program, as Mayr conceives it, is the "proximate" cause of the organism's "goal-directed behavior" (i.e. it explains how that behavior came about in the organism's lifetime, via developmental and other processes). DNA, as an evolutionarily coded set of instructions, programs for, and is the proximate cause of organisms' behavior. A program is not only a sort of blueprint, but also contains what you might call "meta-instructions," that is, instructions on how to make use of the blueprint. Of course, the program itself comes to be through evolution via natural selection over many generations – the behavior's "ultimate" cause.[20] (The proximate-ultimate distinction is useful for pinpointing the relevant *explananda* of teleologically articulated questions or explanations; keeping this in mind will be useful in this, and the following, sections).[21]

A fourth category is purposeful behavior. This is the category with which we may be most familiar in the case of human beings - purposeful acts that follow from planning. The reason this category is not limited to humans is because we see that many "higher animals" (e.g. some warm blooded vertebrates) can engage in these processes, exhibiting a capacity to learn, remember, and plan (Mayr cites as examples some species of birds which hide and retrieve seeds at different seasons).[22] A subset of purposeful behavior might apply *only* to humans - this is the sub-category of planned behavior which we take for granted in our own lives: purposeful acts that follow from our intentions; we consciously deliberate, form strategies,

[18] Take care to note Mayr's particular sense of "teleonomic," in contrast to others previously mentioned. See footnote 12.
[19] Note that the term "program" has fallen out of fashion, perhaps due to concerns that it can be taken to imply genetic determinism (which ignores the contributions of development and the environment towards the phenotype). Mayr, "The Multiple Meanings of 'Teleological,'" 36; Mayr, "The idea of teleology."
[20] Ernst Mayr, "Cause and effect in biology," *Science* 134, no. 3489 (1961): 1501-1506; Niko Tinbergen, "On Aims and Methods of Ethology," *Zeitschrift für Tierpsychologie*, 20 (1963): 410–433
[21] The use of "ultimate" in this context should not be confused with the notion of an "ultimate explanation," meant to reference something like, *that on which all else existentially depends*, e.g. God as the ultimate grounding of, or explanation for all that exists. See Richard Swinburne, *Is There a God?* (Oxford: Oxford University Press, 1996).
[22] Although, as with many cases, we should not be so confident that we know exactly what is going on. This example highlights an ongoing area of interest e.g. Piero Amodio et al., "Testing two competing hypotheses for Eurasian jays' caching for the future," *Scientific Reports* 11, no. 1 (2021): 1-15.

and carry out plans toward chosen ends.[23]

The final category is adapted features. Mayr's categories are based on usage in the biological literature, so we should note that this category is descriptive and not prescriptive. That is, some folks have and may currently refer to adapted features of organisms as teleological. But Mayr thinks it is misleading to do so. Natural selection itself is not goal oriented. Natural selection will reward the production of successful variants i.e. traits or genes that confer fitness advantages upon their bearer. Whether some trait confers an evolutionary advantage can only be determined after the fact. The same is true with subsequent changes of that trait. More simply, "teleological" is a term meant to apply to processes. Either one misapplies "teleological" to a feature which is not *a process* in the relevant sense (e.g. a physical structure), or one applies it to the larger process of evolution (which is inappropriate for reasons previously given). To describe adaptive features as teleological blurs the two in an incoherent, and unhelpful, manner.

To the reader who finds themselves asking: *ok,* but *isn't a program an adapted feature?* The answer is yes. However, what is described as teleonomic is not the feature itself, but a goal-driven process enacted or directed by that program, such as gazelles stotting, bacteria progressing along a chemical gradient, birds preening, etc. To think of adapted features as themselves being teleological may import notions of evolution as progress or the inheritance of acquired characteristics, or promote a misconception that evolution is "forward looking" (as if novel mutations and their associated traits arise *because* populations require them to progress in some way). That is why I think the relevant point to make is not actually about "adapted features," as Mayr dubs this class of usages, but rather that the process of evolution by natural selection itself (the process that brought about such features) is not teleological. The features are not adapted in the sense that they have completed an intergenerational quest of perfection.

Before moving on, I might use this latter point to prefigure later portions of the thesis by saying that while I agree with Mayr that it is confusing to speak of features themselves as

[23] Mayr, "The idea of teleology," 123. In Earlier work, Mayr defends the utility of teleological language in biology but still prefers to avoid anthropomorphism. He was therefore only happy to speak of "goal-directed activities" in the descriptive sense, but chose to avoid talk of "purpose" and "intention." Later, Mayr admitted talk about purposive behavior and planning, but not intentions.

teleological, it will still be useful to use *teleological language* when providing explanations about such features (and I believe Mayr would agree with this). In other words, saying something is teleological is not the same as employing teleological language to explain something (whether that something is rightly or not called teleological). I will develop the case for this kind of heuristic use throughout this chapter and the next. But first, let us attend to a topic that naturally fills in the gap between this section and the next: "functions."

4.2.2 Biological Functions

Teleological language is often used when talking about biological function. In prior work (i.e. prior to Mayr's work), Beckner articulated his own categories of teleology: function, goal, and intention.[24] We have yet to discuss the first of these, and it needs unpacking because Mayr notes two ways "function" is used in biology: pertaining to physiological functioning or pertaining to biological roles of functions.

The notion of biological function carries with it a sense of purpose and explanation. What is the function of a kidney? To filter the blood and remove waste. Filtering the blood and removing waste is what it does and why it is there. But suppose I ask the question differently: why do humans have kidneys? We can give a short-term explanation about genetics and ontogeny: a kidney is the eventuality of various genetically-driven developmental processes (or as Mayr would say, a "program"). Alternatively, we can give a long-term explanation invoking natural selection: we explain the structure and function of the kidney in terms of gradual evolutionary change and accumulative adaptations (or "refinements" of an existing adaptation).[25] Earlier I introduced Mayr's label for these two types of causes or explanations as "proximate" and "ultimate," respectively. My mundane craving for sweet foods can be said to have both kinds of causes and be explained in both ways. Proximately, I see a candy bar, feel hungry, and desire to satisfy that hunger.[26] Ultimately, my ancestors that were inclined to take every opportunity to ingest calorie-dense foods better survived to reproduce

[24] Morton Beckner, *The biological way of thought* (University of California Press, 1959).
[25] Of course, atomizing the kidney in our explanations is too neat, but we can still point to adaptations of whole systems and body types that have been refined over millennia.
[26] In many cases we will want more a fine-grained, mechanical description e.g. ghrelin hormones activate neurons in my hypothalamus to stimulate the release of neuropeptide Y and Agouti-related proteins.

and pass on that inclination such that their descendants, myself being one of them, is endowed with a genetic predisposition to crave sugary treats.[27]

There are several views in the philosophical literature on how to make exact sense of biological function claims.[28] To say something has a function is to neatly condense an otherwise complicated causal explanation – but an explanation of exactly what type? Sometimes the explanation of a function focuses on the role the function plays in the larger system of which it is a part.[29] Other times explanations emphasize how the effect of a function explains its existence.[30] The latter is arguably more useful for questions about evolved functions. For example, one view says that the function of a heart is whatever it was selected for (the "selected effects" account of biological function).[31] We can pinpoint with precision the function of a heart by identifying its purpose: to circulate blood. We can rule out that it was selected to make thumping sounds (an effect, but not the function).[32]

Mayr says our "why" questions are ambiguous. They might mean "how come?" or they might mean "what for?"[33] Unlike some others, Mayr reserves "function" for the physiological description of a heart pumping blood and not the biological role the heart plays, i.e. why the heart is there to pump blood in the first place. So, for Mayr, "function" in the former sense is what it is there for regardless of how it has come about. The latter "why" is about adaptedness, due to evolutionary causes. We could then, adopt for the latter, the available term "proper function."[34] Karen Neander says proper functions "are effects for which traits were selected by natural selection."[35] The proper function of a trait is the function performed by that trait's ancestral trait that contributed to the survival and fitness of its bearer, and in this way ensured its present existence in descendent bearers (i.e. bearers of

[27] This example is commonly used to demonstrate that adaptations may function to bring about effects that later confer no advantage, e.g. in a new and different environment (i.e. there is certainly no shortage of sweet treats on offer today in affluent countries).
[28] For a succinct review of the literature on this topic, see Karen Neander's "Biological Function" in *Routledge Encyclopedia of Philosophy* (Taylor and Francis, 2012), https://www.rep.routledge.com/articles/thematic/biological-function/v-1.
[29] Robert Cummins, "Functional Analysis," *The Journal of Philosophy* 72, no. 20 (1975): 741-65.
[30] Larry Wright, "Functions," *Philosophical Review* 82 (1973): 139-168.
[31] Karen Neander, "Functions as selected effects: The conceptual analyst's defense," *Philosophy of science* 58, no. 2 (1991): 168-184.
[32] A very common example, e.g., Peter Godfrey-Smith, *Complexity and the Function of Mind in Nature* (Cambridge: Cambridge University Press, 1998), 15.
[33] Mayr, "Cause and effect in biology."
[34] Ruth Millikan, "In defense of proper functions," *Philosophy of science* 56, no. 2 (1989): 288-302.
[35] Neander, "Functions as selected effects," 168.

the extant trait in question, e.g. humans with gallbladders that store bile). An ultimate explanation of a trait's proper function answers our "why" questions. An artificial heart can have a function, but by definition, not a proper function.[36] For many cases, the distinction is merely going to be one of emphasis and the nuance between the two senses will be unimportant: in the case of the heart pumping blood, the function and proper function are the same – the function of the heart is to pump blood.

There is a close relationship between mechanical versus proper functions and proximate versus ultimate causes. Proximate causes and mechanical explanations of function are often presented in the context of an immediate response to some other processes (e.g. perturbations from the environment or nerve signals). Ultimate causes and explanations of proper functions involve remote, long-term evolutionary pressures.[37]

4.2.3 How I Learned to Stop Worrying and Love Teleological Talk

I think we can stop here without getting further into the weeds on the topic of biological function. Let us return to the multiple understandings of "teleological." One may think that Mayr's taxonomy (teleomatic, teleonomic, adapted features, purposive behavior, and cosmic teleology) preserves more nuance than is necessary for distinguishing between real and "as if" teleology. It would be desirable to have a simple pair matching the motivations for, and justifications of, our usage: literal and heuristic.[38] For the purposes of my own account, I am happy to proceed with simply, "teleology" and "as-if teleology," emphasizing the heuristic use of the latter. However, I have little need to engage with these terms directly. My interest is in one specific kind of teleological language: agential language (featured in the next section). So, the *actual versus as-if* distinction is what matters, and therefore any language that

[36] Unless you want to make the case that "proper function" can be applied to systems undergoing forms of selection other than natural selection.
[37] In more recent writing on the topic David Haig has argued against running these pair together. David Haig, "Proximate and ultimate causes: how come? and what for?" *Biology & Philosophy* 28, no. 5 (2013): 781-786.
[38] Reese's "Ontological" versus "epistemological" teleology may seem more appealing (Reese, "Teleology and teleonomy in behavior analysis," 78). However, "ontological teleology" refers specifically to "Cosmic Teleology." So, this distinction would not help us in cases in which we want to distinguish between intentional versus non-intentional behavior. How about just using "intentional" and "non-intentional" then? In this case, "non-intentional" introduces further ambiguity (e.g. it could apply to just three or all five of Mayr's categories). I suspect for many cases, the relevant distinction to make is in regards to planned versus programmed goal-directedness, and so perhaps the distinction of teleological versus teleonomic will do. But, of course, Mayr would not agree with the latter term used in this way (i.e. for "teleonomic" to stand in for multiple categories of "not-teleological"). Even without Mayr's taxonomy, we should be wary of using this term due to various historical usages (which recall is what motivated the present analysis).

preserves this distinction will be suitable when describing agents and their processes. But this distinction can be made more precise (i.e. more precise than "planned" versus "programmed" goal-directedness) by drawing from others who have made similar distinctions. For example, consider Reece's suggested distinction between purposefulness and purposiveness.

> Purposive means serving a purpose, and purposeful means having a purpose; purposive means attaining some end or goal, and purposeful means having some end or goal… adaptive behavior is necessarily purposive but not necessarily purposeful.[39]

This distinction neatly delineates the two sub-categories I have indicated within Mayr's single category of "purposeful behavior." Purposeful action is deliberate and follows from mental actions unique (so far as we know) to human agents. (In that vein, consider that many philosophers of action articulate in various ways taking ownership of one's chosen action as constitutive of agency).[40]

As we have seen, useful distinctions will not be limited to behaviors. We can similarly adopt parsimonious pairs in terminology to distinguish various structures, products, or artifacts. Richard Dawkins offers the term "archeo-purpose" when referring to the purpose of a bird's tail (i.e. adaptive functionality), as opposed to the purpose of a plane's tail (with "neo-purpose" serving as the contrasting term for "the kind of 'purpose' we are all familiar with from our own designs and schemes and goals").[41]

Other taxonomies might also be useful so long as we remain careful to separate the useful senses and descriptions from the ambiguous vocabulary they have been given. Outside of biology in the field of cybernetics, Bigelow and Rosenblueth take a different approach. They use the term "teleology" broadly so as to apply to a number of living beings and artificial systems (i.e. machines).[42] For these cyberneticists, teleological behavior is purposeful ("purposive" in Reese's sense)[43] behavior informed by (negative) feedback, defined as when

---

[39] Reese, "Teleology and Teleonomy in Behavior Analysis," 75-76.
[40] E.g. Harry Frankfurt, Christine Korsgaard.
[41] Richard Dawkins Foundation for Reason & Science. "The Purpose of Purpose - Richard Dawkins." Online video clip. *YouTube*, 19 June 2009. Web. 1 October 2019.
[42] Rosenblueth et al., "Behavior, Purpose and Teleology," *Philosophy of Science* 10.1 (1943): 18-24.
[43] "The term purposeful is meant to denote that the act or behavior may be interpreted as directed to the attainment of a goal - i.e., to a final condition in which the behaving object reaches a definite correlation in time or in space with respect to another object or event. Purposeless behavior then is that which is not interpreted as directed to a goal." (18). They give as an example a target-seeking torpedo as "intrinsically purposeful." While

"the signals from the goal are used to restrict outputs which would otherwise go beyond the goal."[44] While this sense of "teleology" is less helpful in carefully disambiguating intentional activity from programmed or non-intentional adaptive behavior, if read as *as-if teleology*, it does represent a broader interest in the project of "naturalizing" teleology shared by many in the biological literature (e.g. Millikan), and importantly, accommodates a wider variety of systems for which we may deem teleological language appropriate or useful in our theorizing and explanations (with self-reflective human beings acting from intentions constituting a special subset). I think we can take a similar approach generally to other accounts and usages of teleological language.

The pay-off for this more coarse-grained distinction ("actual" versus "as-if") will be that many claims will depend on one or the other, but we can make sense of them in the same way. There is an over-arching understanding of "agency" (i.e. the concept of agency) sufficient for some purposes (e.g. biology) that will encapsulate both the *actual* and the *as-if* flavors.[45] However, this is only plausible after having rid the *as-if* of its more untenable associated notions. We have done this for "teleology" more generally in this section and will do so for "agency" in the following sections. (If it seems somewhat mysterious that I have spent time developing a two-part distinction only to highlight the utility in blurring the difference, I hope to make this clear in the following sections. Briefly, whether or not one is really an agent will matter a great deal in the domain of moral and political philosophy; but when the concept of agency is used heuristically in biological theorizing, that question is nearly irrelevant).[46]

In conclusion, the terminology discussed in this chapter is a means to an end: to disentangle the various ways "teleology" has been used and understood in the past. We can rid ourselves of some of the more dubious notions associated with a classical understanding of teleology. We can also discard the over-prescriptive "teleological taboo" which would disallow *all* teleological language in biology. In our discussion of programs, adaptedness, and function,

---

space does not allow for further development in this thesis, the concept and role of interpretation should be further investigated comparing the behavioral versus organizational accounts in cybernetics with the and interpretivist versus functional accounts of group agency.

[44] Rosenblueth et. al. "Behavior, purpose and teleology," 19.

[45] This should not be read as saying anything specific about particular distinctions. For example, in Reece's sense, something can be purposeful without being purposive, so in this case the *actual* is not simply a special case of *as-if*, i.e. *as-if* with further capacities or conditions met.

[46] Especially when using models of actual rational agents that are, quite ironically, not much like real human agents (e.g. "homo economicus"). See section 4.5.

we have seen that teleological language can be naturalized and deployed unproblematically. We can confidently conclude that the judicious use of teleological language in biology is both justified and useful.

4.3 Agential Thinking

In this section I will further the case for a heuristic use of a specific type of teleological biological reasoning: that of agents with goals.

We have seen that explaining a behavior or trait by its effect need not imply backward causation. In light of natural selection, biologists will often invoke the language of design, function, and purpose. Samir Okasha argues that we may with legitimacy also invoke the language of intentional agency, heuristically conceiving of an organism as a rational agent.[47] Accordingly, we may evaluate and understand an organism's traits or behaviors as though they were strategies deployed in pursuit of a goal. The goal of a biological agent is survival and reproduction. Talk of strategies and goals naturally invites the accompanying language of what an organism "tries" to do, reacting to its environment as if it "knows" to do this or that in light of what it "wants" to achieve (in philosophical terms, practical reasoning and instrumental rationality).[48]

4.3.1 Adaptationist Reasoning

Okasha classifies the above way of "agential thinking" as a "form of adaptationist reasoning" – a way of "trying to understand evolved traits in terms of their contribution to fitness."[49] An adaptation has been defined as "a feature that has become prevalent in a population because of a selective advantage conveyed by that feature in the improvement in some function."[50] Sometimes adaptations are defined generally as any characteristic of an organism that adds to its fitness,[51] but most other definitions include an explicit requirement that

[47] This is a further step than Mayr, who resisted anthropomorphisms.
[48] Samir Okasha, *Agents and goals in evolution* (Oxford University Press, 2018), 13.
[49] Okasha, *Agents and goals in evolution*, 2.
[50] Douglas Futuyma, *Evolution*. 2nd ed. (Sunderland: Sinauer Associates, 2009), G-1.
[51] Ernst Mayr, *What Evolution Is* (New York: Basic Books 2001), 283.

adaptations must come about via a certain process (i.e. natural selection, see "adaptationism" below).[52]

So, the term is used to indicate both the process of the population adapting to the environment through trait modifications, and the outputs of that process, i.e. the adaptive traits. The term "adaptedness" refers to the state of organisms or traits at the end of this process, having suitably adapted to their current environment.[53]

Lastly, "adaptationism" can refer to a research program, a way of reasoning, or a position on what role natural selection plays and how we might or should proceed on that view.

Peter Godfrey-Smith distinguishes three senses of "adaptationism." The first kind of adaptationism ("empirical adaptationism") is the position that natural selection is ubiquitous and "causally pre-eminent."[54] Natural selection is sufficient to make predictions and provide explanations about the vast majority of the traits we see on display in the living world. A second sense of adaptationism ("explanatory adaptationism") is the position that natural selection is the "big answer" to the "big questions" in biology – questions we have about the "apparent design of organisms and the relations of adaptedness between organisms and their environment."[55] Unlike the former sense, the latter is not an empirical claim. The third and final sense of adaptationism ("methodological adaptationism") is a prescription for doing science in this area: "the best way for scientists to approach biological systems is to look for features of adaptation and good [apparent] design."[56]

In summary, "adaptationism" is used variously to say that as a matter of fact, natural selection is causally powerful and important; that natural selection is the best way to explain adaptedness (regardless of how rarely it acts or how important it is); and, that evolutionary biologists should organize their research around features of adaptation. Whereas this first

[52] Nicholas Barton et al., *Evolution* (Cold Spring Harbor: Laboratory press, 2007), 10.
Carl Bergstrom and Lee Alan Dugatkin, *Evolution,* 1st ed. (New York: Norton, 2012.), G-1.
[53] Of course, there is no real "end" and what we may count as adaptedness can change with e.g., the frequency of traits in the population or environmental fluctuation.
[54] Peter Godfrey-Smith, "Three kinds of adaptationism," in *Adaptationism and Optimality*, eds., S. Orzack and E. Sober (Cambridge University Press, 2001), 338.
[55] Godfrey-Smith, "Three kinds of adaptationism," 336.
[56] Godfrey-Smith, "Three kinds of adaptationism," 337.

sense is a kind of realism about adaptationism, the second and third are normative (evaluative and prescriptive, respectively).

All three senses may be relevant when arguing for the heuristic value of agential thinking, but the second and third kinds, "explanatory adaptationism" and "methodological adaptationism," are what Okasha has in mind when he says that agential thinking is adaptationist reasoning. Natural selection is responsible for the unity exhibited by living organisms – the unity that warrants agential attributions (see section 4.4.1). The value of methodological adaptationism is most obvious when we observe features that are maladaptive to the focal entity. As I will discuss in chapter 5, such instances will show us that it is sometimes not adaptationist reasoning that is at fault, but rather how we have applied it.[57] Indexing traits to the goals of agents allows us to make sense of what explanatory adaptationism (seemingly) cannot, as what may be maladaptive for the focal individual may be adaptive for another (this will be developed further in section 5. 2). Godfrey-Smith's three categories (especially the third) will make a reappearance in section 5.2.6, in which I adapt them for more specific considerations about agential heuristics.

For the purposes of this thesis, I will set to the side more nuanced debates about the legitimacy of each of these forms of adaptationism. It is the *distinction* among the three kinds, not the ultimate acceptability of these positions, that matters to my discussion later. However, I will linger to mention just one critique of adaptationism more generally.

Gould and Lewontin are skeptical of the adaptationist program, which they see as a combination of the atomization of traits plus the presumed optimization of parts via selection.[58] This skepticism is born of the after-the-fact nature of identifying adaptations. They see adaptationist claims as largely unfalsifiable with the heuristic approach many feel to be valuable as inviting the telling of "just-so" stories, with the criteria for acceptance of such

[57] Using adaptationist reasoning may be more helpful with some things than others. Godfrey-Smith considers whether the prescriptive adaptationist approach he discusses (i.e. "methodological adaptationism") might be adopted selectively. That is, perhaps deploying agential heuristics will be more valuable with some traits and behaviors over others, or at some levels of description over others, or with some types of focal entities over others. I do not consider these possibilities in this thesis; however, I do note it as a potential source of objections to Okasha's work (and my own) in which it is of crucial importance that agential thinking is applied the same way at different levels of agency and individuality.

[58] Gould, Stephen Jay, and Richard C. Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm: a critique of the adaptationist programme," *Conceptual issues in evolutionary biology* 2 (1994): 73-90.

stories being mere consistency with natural selection.[59] Gould and Lewontin charge adaptationists with atomizing organisms into traits, which are "explained as structures optimally designed by natural selection for their functions."[60] In contrast, they argue, we should see that organisms are "integrated entities, not collections of discrete objects."[61]

But in response to this very point, Okasha responds that those who employ the "as-if" agential heuristic are clearly aware that "attributions only apply when organisms… have traits which contribute to a single overall goal."[62] Okasha's brand of adaptationism is a presupposition, not dismissal, of organismic unity. So then, Gould and Lewontin's point about integration and unity is consistent with, and in fact central to, Okasha's criteria for agential attributions. This will become clear in the next section.

4.3.2 Agents and Their Goals, Traits and Their Functions

So then, agential thinking is a form of adaptationist reasoning. To reason like an adaptationist is to first observe evolved traits, and second, to consider those traits in light of the effects they may have on fitness. Organisms will bear traits or behaviors that benefit them in terms of fitness (the survivors best suited to their environment have survived and flourished, and their adapted traits have become widespread or fixed in the population). Organisms bear adaptations which, to the extent that they are unconstrained to do so (e.g. by phyletic or architectural constraints), maximize fitness. In other words, these traits confer more advantages relative to other non-adaptive traits, and thus we can see them as evolutionarily honed strategies. Agential thinking draws on a fruitful parallel between "fitness" in evolutionary biology and "utility" in economics.[63] Therefore, we reason as if an organism's adaptive traits and behaviors were its choices, i.e. chosen for their utility from among a range of alternative candidate traits and behaviors.

"Agential thinking" is not a new idea. Okasha's work on the subject is an observation, analysis, and emphasis on the utility of agential thinking. The phrase itself originated with

[59] E.g., Gould and Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm," 86.
[60] Gould and Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm," 83.
[61] Gould and Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm," 83.
[62] Okasha, *Agents and Goals in Evolution*, 34.
[63] John Maynard Smith, *Evolution and the Theory of Games* (New York: Cambridge university press), 1982; Alan Grafen, "Optimization of inclusive fitness," *Journal of theoretical Biology* 238, no. 3 (2006): 541-563; Alan Grafen, "The formal Darwinism project in outline." *Biology & Philosophy* 29, no. 2 (2014): 155-174.

Peter Godfrey Smith.[64] Even earlier, Robert Wilson dubs this mode of thinking "the cognitive metaphor,"[65] and earlier still, Elliot Sober wrote about "the heuristic of personification."[66] Most famously, Richard Dawkins asked his readers to take the perspective of a gene that acts "selfishly" in pursuit of replicating itself.[67]

Most intuitively, we take up this agential thinking when considering multi-cellular organisms that look and behave somewhat similarly to humans, for example when we observe the foraging patterns of birds and draw conclusions about their "intentions," "methods" and "planning." But the notion of biological agency may likewise be projected onto single-celled organisms, groups of organisms or sub-organismic elements, i.e. a colony or a chromosome. I follow Okasha in applying the heuristic in this way.[68]

On the minimal notion of an agent in which an "action" is some behavior (i.e. *doing something*) attributable to an internal cause within that entity,[69] the world is full of agents at many levels: genes, cells, plants, fish, swarms. On the sophisticated philosophical notion of actions coming about through intentions formed by desires and beliefs, perhaps only humans (as far as we know in fact, but perhaps also aliens and angels in theory) make the grade. But the utility of agential thinking is not constrained by what we *really* take to be agents. Agential thinking involves treating biological entities as "agent-like, not as literal agents."[70] Thus, consistent with the "as-if teleological" sense of "teleonomy," agential thinking invokes a similar "as-if" understanding of biological agency.

The functional idiom applies to traits; traits have functions. The agential idiom presumes a unity among traits; an organism's traits function harmoniously to achieve the agent's goals.[71] It is possible to identify the function of some trait that would have us repeal such a

---

[64] Godfrey-Smith, *Darwinian Populations and Natural Selection*, 10.
[65] Wilson, *Genes and the agents of life.*
[66] Elliot Sober, "Three differences between deliberation," in *Modeling Rationality, Morality, and Evolution*, ed., Peter Danielson (Cary: Oxford University Press, 1998), 408.
[67] Richard Dawkins, *The Selfish Gene. 40th Anniversary ed*. (Oxford: Landmark Science. 2016).
[68] Applying agential thinking to natural selection itself, however, would be a mistake (Okasha argues this point with Dan Dennett as his foil, who portrays mother nature as an agent who selects optimal variants). While the heuristic can be profitably applied to products of evolution, it should not be applied to the process. Dennett, *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (New York: Simon & Schuster, 1995); Dennett, *The Intentional Stance*.
[69] Fred Dretske, *Explaining Behavior: Reasons in a World of Causes* (Cambridge: MIT Press, 1988).
[70] Okasha, Agents and Goals, 14.
[71] Okasha, Agents and Goals, 30.

presupposition, such that we then deem use of the agential idiom inappropriate. In such a case, each trait can still be described in terms of function, but it must not be said to function *for the agent.*

To explain, I will use Okasha's preferred example of cytoplasmic male sterility in flowering plants.[72] It can sometimes be observed that an otherwise hermaphroditic plant lacks a stamen or has a stamen that fails to produce pollen. How can such a trait be thought to be conducive to the plant's goal? How can sterility be viewed as a strategy to increase plant fitness? It cannot. Sterility is caused by mitochondrial genes, which are matrilineal and therefore have no "interest" in pollen. The trait of male sterility has a function, but that function serves mitochondria. *Functional* talk makes perfect sense here, but agential talk does not. In chapter 5, I will develop this example of traits being at odds with each other (nuclear interest in stamen production versus the mitochondrial interest in foregoing pollen production) as an example of disunity - that is, insufficient unity to meet the criteria of agency or even warrant agential attributions ("as-if" agency).

I think it is important to emphasize the difference between "agential thinking" and the use of agency concepts in biology more generally. I will argue in section 4.4.1 that Okasha's "unity of purpose" criterion makes as-if-agency in biology sufficiently analogous to human agency in philosophy such that rational and biological unity can function as shared concepts.

Before moving on, it is worth briefly mentioning two other philosophers who also make use of the connection between agents and biological individuals, and explaining why I have not chosen to feature them in this thesis.

In *Organisms, Agency, and Evolution,* Denis Walsh sets out to emphasize the ways in which organisms can be seen as participating in and shaping their own evolution.[73] It is in this sense, that Walsh sees organisms as agents. While Walsh's work rounds out a picture of evolution that is most often left incomplete by reductionist treatments (e.g. gene-centered accounts), his sense of agency is different from my own.

Robert Wilson is another philosopher who discusses agency in a biological context in his book, *Genes and the Agents of Life.* While Wilson provides very useful distinctions and categories for

[72] Okasha, Agents and Goals, 30.
[73] Denis Walsh, *Organisms, agency, and evolution* (Cambridge University Press, 2015).

living and non-living parts, individuals, and collections, his use of "agent" applies more widely than suits my purposes in this thesis.

In contrast to these two philosophers, my argument for the utility of agential heuristics in biology is independent of whether and which biological entities *really* are agents. It is sufficient that the entities towards which we can take up the intentional stance by extrapolating from their rational point of view have *at least* "as-if" agency. Given space constraints and my particular interest in "as-if" agency, I have chosen to privilege Okasha's account.

4.4. Rational Unity

In this section I will argue that "rational unity" is what warrants agential attributions for actual and as-if agency alike.

4.4.1 Unity

The topics of biological individuality and philosophical agency share the common theme and criteria of unity (as surveyed in chapters 2 and 3). An agent is not a series of unassociated intentional episodes. An agent is that in which such experiences and mental states are meaningfully integrated so as to meet the basic conditions of consistency and transitivity. A biological individual counts as such because of similar coordination and integration of its parts that share a common fate. An individual could not have come about otherwise. Thinking and speaking about biological entities as if they were agents is justified in cases when they exhibit an alignment of evolutionary interests and a co-functioning of components – what Okasha calls a "unity of purpose." The use of agential heuristics is warranted when the parts of a biological entity work in concert towards the "goal" of the whole.

Our conceptual common currency then is unity. Using the sisterly concepts of "agency" and "as-if agency" we can more specifically trade on the concept of "rational unity." We might, in the spirit of Rovane and Korsgaard (section 2.2), describe an agent as deliberating and acting from a rational point of view.[74] An agent coordinates its actions by making decisions about

[74] Rovane and Korsgaard will mean *deliberating* in a literal way, as people do. What our "as-if" sense of agency is borrowing is the "rational point of view" bit.

what is in its interest. For example, the members of a group agent *reason* together so as to achieve this unity through deliberation and produce a group-level output; group level rationality then is an *achievement.* We do not need to posit that biological entities reason, but rather that evolutionary interests are aligned, such that "strategies" (traits) are coordinated according to a shared "goal" (of maximizing fitness of the whole). The characteristic unity exhibited by biological organisms is due to the cooperative functioning of their parts. Unity of purpose is an *achievement* of integration and collaboration. Individuality is an *achievement* of coadaptation.[75] When attitudes of an agent or traits of an organism are not sufficiently integrated to achieve a unity of purpose, our attributions of agency and individuality are misguided. The normative component of rationality should be understood in terms of this achievement of unity (i.e. without which one "fails" to warrant agential attributions).

4.4.2 Rationality

Economic models of rationality are based on assumptions about rational agents behaving to maximize utility. If given a choice between two alternatives, the agent will choose that which best satisfies the agent's preferences. A rational agent is consistent in holding and acting on her preferences. An agent's preferences can be ordered and satisfy conditions of transitivity. A "preference" should be understood broadly to include things like a preference for paying taxes, which itself might simply be a preference to participate in the system and avoid penalties. This latter example hints towards the more interesting cases in which outcomes are uncertain and an agent must choose based on probabilities and expected outcomes.

Non-intentional subjects (e.g. protists) can "choose" via behavior in the economic sense (e.g., singled-celled organisms following a nutrient gradient). We can observe many types of non-human agents that fulfill standards of rationality and act to maximize expected utility without ever expecting them to be capable of holding reasons for their actions.[76] (Dennett makes a similar point regarding the intentional stance: whether or not an organism is "said to *believe* the *truths* of logic, it must be supposed to *follow* the *rules* of logic").[77]

---

[75] This is not to imply every component is an adaptation. Some may not be, and others may not even be proper candidate traits but rather artefacts of our clumsy parsing of features, e.g. chins and elbows. Also, maybe some will want to use the term "co-adapted" in cases of co-evolution among non-integrated entities, i.e. butterflies and the larger fauna that their wing patterns have evolved to resemble.
[76] Okasha, *Agents and Goals*, 152-153.
[77] Daniel Dennett, "Intentional systems," *The Journal of Philosophy* 68, no. 4 (1971): 95.

Following Kacelnik, we might refer to E-rationality and P-rationality to separate the economist's and philosopher's senses of rationality, respectively.[78] Kacelnik also uses the term "B-rationality" for the evolutionary sense of maximizing fitness – something is B-rational if it is adaptive. However, in the context of this thesis, we can collapse the categories of E-rationality and B-rationality, having adopted the agential heuristics outlined above. However, in doing so we should note a distinction between the two. B-Rationality is an externalist concept: "fitness consequences of a given behavior depend on the environment" – so B is a special case of E.[79]

The distinction between E-rationality and P-rationality might also be collapsed if we take the former to simply be a formalization of the latter. However, I will proceed by taking the position that while it can be the case that E-rationality is sometimes a mere formalization of the practical rationality of human agents, it need not always and only be. After all, in rational choice theory we make assumptions about human beings we know are false (e.g. completely selfish), but the models still prove useful in making predictions about human behavior. Also, when and if rational choice theory proves unhelpful for human actors (e.g. when a sense "fairness" confounds our assumptions about what choices agents will make), it might still prove quite useful for as-if agents (e.g. biological entities "acting" to maximize fitness). So then, we can sometimes make sense of and make predictions about the behavior of non-human agents using economic rational choice theory just as we do with human agents. It does not follow that we should either ascribe reasons and beliefs to all agents exhibiting E-rationality or, on the other hand, that we should reject the application of rational choice models to non-human agents all together.[80]

For our purposes the relevant sense of rationality is the economic sense as used in rational choice theory. This sense captures two important ideas from chapters 1 and 2: agency entails an inherent point of view from which an agent deliberates and acts (as in the practical rationality emphasized by Korsgaard); and, to count as an agent means to satisfy conditions of rational unity, e.g. logical consistency (as in the theoretical rationality emphasized by LP).

---

[78] Alex Kacelnik, "Meanings of rationality," in *Rational Animals?* eds., S.E. Hurley and M.E. Nudds (Oxford: Oxford University Press, 2006) 87–106.
[79] Okasha, *Agents and Goals*, 153.
[80] A minor point for potential future development: perhaps it would be more useful to talk of Behavioral Rationality and Intentional Rationality e.g. B-Rationality, I-Rationality.

4.5 Parallels in Methodology – Shared Commitments and Similar Problems

In this short section I will briefly highlight some parallels in thinking about groups and individuals across domains.

Okasha draws a parallel between methodological individualism and adaptationist modes of understanding. Both the ideas of sociologist Max Weber[81] and adaptationist G.C. Williams[82] are commitments to a kind of explanation that prioritizes the individual. For Weber, only individuals are agents whose actions are intelligible through their intentional actions, and therefore scientific explanations should be given in terms of individuals and their choices, not groups. For Williams, it is individual organisms that exhibit adaptations, and therefore the traits that evolve should be explained based on the survival and reproductive advantages they confer on individuals, not groups. Okasha argues that the logic of both arguments is the same: "a commitment to a particular mode of explanation, plus a claim that the property required by this mode is only possessed by individuals."[83]

But of course, LP (and others) aim to show how groups can be unified actors. A group agent is *an* agent. The inherent ambiguity regarding the terms "agent" and "individual" has been addressed in chapters 2 and 3 (with further critique to follow in chapter 5). Both individuals and groups have intentional episodes, make decisions, and act on preferences. Individuals and colonies can bear adaptations, exhibiting as-if agency in pursuit of advancing their proximate and ultimate goals. While Okasha's point is to draw a parallel between economic and evolutionary modes of explanation such that we might fruitfully and justifiably borrow and loan concepts from each field, we should further note the parallel definitional and taxonomic ambiguity as well as the multi-level applications. As we will see in section 5.1, "agent," and even "individual agent," are not useful categories for some of the concerns motivating the work on group agency in moral and political philosophy. Similarly, we will see in section 5. 2 that delineating "individuals," even of different kinds, is not sufficient to capture what is often most interesting or relevant in our explanations of biological individuality.

[81] Max Weber, *Max Weber on Law in Economy and Society* (Cambridge: Harvard University Press, 1954).
[82] G. C. Williams, *Adaptation and Natural Selection (*Princeton: Princeton University Press, 1966).
[83] Okasha, *Agents and Goals,* 62.

So then, we can observe not only parallels in the commitments to certain methodologies between these two literatures, but parallels in the kinds of problems or shortcomings that may result (i.e., what is left wanting from the concepts "individual" and "agent"). These parallels make plausible the hope that consideration of a problem in one domain may help inspire inroads with a similar problem in the other. (For example, in section 5.2.3 I will show how concerns about the potential for disunity in group agents helps reframe our conceptualization of biological individuals). If we are justified in fluidly considering and cross-examining e.g., LP's group agents and Godfrey-Smith's collective reproducers alongside Korsgaard's human agency and Okasha's as-if agency, we might find novel pathways towards a better understanding of both bodies of literature.

4.6 The Unity of Agency

In conclusion of this chapter, we have seen that the concept of rational unity may justifiably be used across domains, imported and exported freely between philosophy of biology and moral and political philosophy. This is a unity in which traits of group agents and biological individuals alike work in concert towards the goals of whole. In the following chapters, I aim to demonstrate why this is not only justified but potentially profitable for both sub-fields.

Chapter 5
The Literatures on Group Agency and Biological Individuality can be Mutually Informative

The main argument of this thesis is that the topics of biological individuality and group agency can be mutually informative. While there might be a variety of projects potentially worth exploring along these lines, this chapter will give attention to three. (I will briefly sketch some other suggestions in the following chapter). In Section 5.1, I will argue that LP's account (and others like it) could benefit from the pluralistic approach we see in the literature on biological individuality, taking a nuanced approach to "agency" in a way that best captures distinctions that really matter. In Section 5.2, I will argue that the topic of biological individuality could benefit from a multi-level agential heuristic, drawing on a perspective that comes naturally when engaging with topics in the group agency literature. In Section 5.3, I will present a case that suggests the potential for long-term ongoing "conversations" between sub-disciplines.

5.1 Biological Individuality Informing Group Agency

In this sub-section I will begin by highlighting some problems with the categories used by LP in their account of group agency. I will argue that the aspects of their account which make it plausible that groups can be agents and persons have the undesirable effect of categorizing agents in a way that misses making distinct those with interests that matter. I will explain why LP's commitment to normative individualism does not help with this problem. I will suggest that, on their account, terms like "agent," "person", and even "individual" are ambiguous and in need of a more nuanced taxonomy to distinguish different categories relevant to our concerns and undertakings, and most importantly, for identifying what matters to us most.[1] I will suggest options for LP (and for some issues, the group agency literature more generally) based on tools deployed in the literature on biological individuality.

5.1.1 The Insufficiency of "Agency"

[1] *Preserving* what matters most is the practical task which motivates those like LP. *Identifying* what matters most is the theoretical task which must come first.

For LP, agency is multiply realizable. There is no special "agent stuff" required for agency. The criteria that a human must meet in order to qualify for agency are criteria that groups can fulfill. This latter point is what makes group agency plausible. Groups can be agents.

In philosophy, we usually take a paradigm agent to be a healthy, neuro-typical, adult human being. There are two relevant things to say about this paradigm agent:

1. It has morally commanding interests.[2]
2. It is fit to be held responsible.

There are some beings with morally commanding interests that we do not hold responsible for conforming to social norms, and therefore we do not normally consider them to be agents. Examples include infants and perhaps some non-human animals. We maintain that, though not agents, they still have interests that matter. This distinction is sometimes made using the terms "moral agent" versus "moral patient" in which only the former is fit to be held responsible.

What LP have proposed in their account, is that agents can qualify for agency (#2) without having morally commanding interests (#1). That is, what makes an agent fit to be recognized in society, counted as a distinct center of attitude and action, granted certain privileges and held to account, has nothing to do with whether or not that agent has morally commanding interests. In other words, the categories of "moral patient" and "moral agent" can come apart in both directions. It is possible to be either one without the other.

LP aim to convince us that groups can be agents in the sense of being rationally unified, autonomous, moral centers of attitude and action that can be held responsible for their decisions and actions. Though they argue this case persuasively, it comes with an unwanted side effect of (non-trivial) categorical confusion. Groups can be agents, but group agents and human agents are different in *morally* significant ways. Humans have interests that matter, and group agents do not.

[2] This phrase is used by LP (e.g. 182). It does not simply denote moral significance but something that demands a moral response.

Therefore, LP's agency concept lumps together two *kinds* of agents that it is otherwise important to keep distinct for a range of considerations and purposes. LP acknowledge this. Indeed, they state that recognizing such differences is not only central to, but also foundational to their motivations. LP are committed to normative individualism. They are very worried about power asymmetries between, e.g. corporations and individual human persons. They recognize that humans are vulnerable and finite while group agents have potentially unlimited resources and lifespans. LP express their certain conviction that individual agents deserve special protections and that group agents should have the burden of special restrictions. So, in fact, LP are actually calling for a recognition of the difference in moral status between these two types of agents.[3]

The single term "agent," then, seems rather unhelpful. Furthermore, by their own standards, LP cannot justifiably rely on the use of "individual" and "group" to articulate this difference. These descriptions are irrelevant on a multiply realizable account. If agency is multiply realizable, then "individual" and "group," when used in the phrases "individual agent" and "group agent," are mere descriptions of the particular manifestations (i.e. physical instantiations) of two agents. By analogy, if all we care about is whether a clock is a clock, "plastic" and "metal" (or "digital" and "analog," etc.) become irrelevant descriptors of the various clocks we survey.

Therefore, LP's concept of "agency" is unhelpful if we want to ensure that these two kinds of entities are not carelessly lumped together. Human agents have morally commanding interests. Group agents do not. The same factors that allow groups to be agents on LP's account (e.g. multiple realizability) are at odds with providing a justification for their endorsement of normative individualism.

If LP's concept of agency is insufficient to support their own commitment to normative individualism, will their distinction between "persons" and "agents" be of help here? I will argue, no. In fact, LP's account of personhood makes the problem clearer. So then, let us turn to what we might call LP's "personhood problem."

5.1.2 The Personhood Problem

[3] List and Pettit, *Group Agency*, 182.

Recall from chapter 1 that according to LP, it is not only possible that groups can be agents – they can also be persons. As with agents, there is nothing intrinsic to being a person - persons are multiply realizable (with no special "person stuff" required).[4] Persons are those that perform like persons. Persons are agents who perform effectively in the space of obligations. Persons are

> party to a system of accepted conventions… under which one contracts obligations to others and… derives entitlements from the reciprocal obligations of others… to be a knowledgeable and competent party to such a system of obligations… to recognize that censure or sanction are reasonable in cases of failure.[5]

LP make the case that some groups can fulfill these criteria and are therefore persons.

LP cannot at the same time view intrinsicist personhood as a rival concept (that they reject)[6] and *rely* on that concept to discriminate between individual and group "persons." There is a difference between, on the one hand, contrasting two "person" concepts to make an argument for using one over the other, and on the other hand, setting up *two kinds of persons* within the same account.[7] LP engage in the former explicitly (defending their choice of the performative concept), and the latter implicitly (potentially undermining that choice).

This implicit reliance on the intrinsicist concept is seen in LP's commitment to normative individualism. Our history of instrumentally ascribing personhood (legally) and conflating agency and personhood (conceptually) can lead to a significant mistake: not sufficiently regarding the inherent vulnerabilities of individual human persons.[8] It is in virtue of these intrinsic qualities that individuals should enjoy a privileged, protected status. LP's concern for individuals is really a concern for individual *human beings* – sentient, vulnerable agents that matter. While group agents can and should be held responsible as moral actors, they

---

[4] List and Pettit, *Group Agency*, 171.
[5] List and Pettit, *Group Agency*, 173. Note the phrase "reciprocal obligation" and see footnotes 18 and 31.
[6] LP write "The Hobbesian approach generalizes the performative conception of the person beyond the legal context, developing it as a rival to the intrinsicist view" (172).
[7] "The corporation was not a person in …[the] intrinsic sense but only in the extrinsic, performative sense of being able to operate in legal space" (171).
[8] Kendy Hess, ""If You Tickle Us....": How Corporations can be Moral Agents without being Persons." *Journal of Value Inquiry* 47, no. 3 (2013): 319-335.

should not be granted the same moral status as human agents who experience weariness, despair, joy, stress, and suffering. It is only because of these qualities of human beings that agents have obligations in the first place (and as LP argue, this is why group agents should have additional, non-reciprocal obligations towards individual agents).[9]

A history of associating agency with the paradigm case of human beings may lead one to intuitively presume that all moral agents are experiencing subjects. However, group agency, as a concept, allows us to consider a kind of agent that is not a paradigm agent (i.e. a rational human being). We do not assume group agents are subjects of experience (as with some other possible agents, like computer AIs). While group agents operate from a rational point of view, we should not unnecessarily add to this a *phenomenal* point of view (as discussed in section 2.2). LP argue that group agents can perform competently and reliably in the sphere of obligations within society (i.e. "performative personhood"), yet this says nothing about group agents being experiencing subjects. In separate papers, both List and Pettit each make the case themselves that group agents are not phenomenally conscious (though Pettit will allow that groups can possess "access consciousness:" a conscious state interacts with other states, has access to their contents, such that the information is available to the agent).[10] Even if we conclude that computers or group agents *could* be phenomenally conscious, there is no reason to assume phenomenal subjectivity is bound to moral agency.

Why are we talking about phenomenal consciousness? Paradigm agents are phenomenally conscious, rational humans. To experience hope, and fear, and anxiety is to *experience* hope, and fear, and anxiety. Vulnerable human agents can have bad experiences. Humans can feel pain. By contrast, group agents cannot feel pain.[11] Now we are talking about sentience. While some non-human animals, for example, may factor into our moral consideration because of their capacity to suffer, corporations do not, and should not. The only moral qualms we might have about dissolving a corporate entity ("killing" it) would be in regard to the effects that such a dissolution would have on its human members. (Therefore, and to make plain what we are building towards here, we do not regard group agents as moral *patients*, even if we accept

[9] The reader will note the insufficiency of the term "individual" here. This will be developed below.
[10] A "philosophical zombie" could possess access consciousness but not (by definition) phenomenal consciousness. See Ned Block, "On a confusion about a function of consciousness," *Behavioral and brain sciences* 18, no. 2 (1995): 227-247; Christian List, "What is it Like to be a Group Agent?" *Noûs* 52, no. 2 (2018): 295-319; Philip Pettit, "Consciousness Incorporated," *Journal of Social Philosophy* 49(2018): 12-37.
[11] See Rachael Briggs, "The normative standing of Group Agents." *Episteme* 9, no. 3 (2012) 283–291.

that they are moral *agents.* We do not automatically regard all moral agents as moral patients).

It is as least worth mentioning that two things I have naturally run together above are logically distinct: phenomenal consciousness and sentience. We can at least imagine a being that, while an experiencing subject, has no capacity for pain and suffering and does not value their life and future. It is more difficult to imagine one who enjoys unexperienced sentience, but the point remains that the categories might still exist in a world of philosophical zombies. So, a subject of phenomenal consciousness should not be confused with a subject of moral concern (i.e. one who is owed moral consideration). However, the former is often given as the reason for the latter.

LP offer criteria for personhood such that some groups can meet the criteria. LP are also committed to normative individualism, which fundamentally privileges consideration of individuals such that our ethical decisions should be justified with reference to individuals (i.e. as opposed to social classes, political parties, institutions, etc.). On LP's view, groups can be persons, but individuals are the only persons which matter. This is because individuals are sentient and are vulnerable (e.g. prey to dominating and powerful groups). So, according to LP, individuals deserve a special protected status such that they enjoy rights that group persons do not, and group persons have special obligations that individual persons do not. So, on LP's account then, the question of personhood is distinct from the question of what matters. For LP, an agent can be a person without being an agent that matters. (Again, so there is no mystery – the view I am building towards is one that says, contrary to LP's view, "person" should be reserved for the category of agents that matter).

I am developing the criticism that LP should not be ascribing personhood to (some) group agents. But before proceeding I would like to give a slightly broader context for the conversation. There are different views about moral status. According to LP, what makes one worthy of consideration is distinct from what qualifies one for agency. However, one could hold a different kind of view. One could hold the view that it is not sentient beings that matter (as Utilitarians do), but rather rational beings that matter (as Kantians do). Those who hold the latter view will likely not want corporations to be agents because they do not want corporations to have the protected status that individual humans enjoy. That is, they do not want corporations being counted among those who matter. One who holds this view will be

motivated to produce reasons as to why corporations cannot be rational. Yet if groups *can* be rational, and if rationality is what matters, one would need a further argument to defend normative individualism (i.e. in order to exclude group agents from the rights and privileges individual agents enjoy). That argument would presumably need to include what is different about *humans* as opposed to corporations and robots. Further, this view would struggle to make sense out of non-rational beings that we normally think have interests and deserve protection (e.g. infants). (One could, of course, make the case that we are mistaken to think any non-rational beings deserve any consideration, given the view that rationality is what matters for moral status. This would be consistent but highly counterintuitive when it comes to, e.g. babies).

On LP's view, it would not make sense to worry about corporations being agents for the reason that one does not want corporations to matter. Again, this is because LP hold that a group can be an agent without being an agent that matters. Indeed, a group can even be a *person* on LP's view and still not matter. For LP it is individuals that matter. But, as I will argue, "individual" is problematic.

"Normative individualism," while it makes sense prior to and outside of LP's account, becomes a misleading name *within* their account. LP are worried about protecting individual *human beings*. They are *not* simply concerned with individual agents or individual persons. Why do I say this? Because on their view, agents, and the subset of those with agency that are persons, are multiply realizable. Whether they are individual persons or group persons should not matter for this kind of account (no more than it matters for the purposes of telling time (for instance) whether a clock be digital or analog). Any agent that meets the criteria for personhood is *a* person. A group agent is *an* agent. Therefore, LP use the term "individual" within the context of an account wherein this word only picks out irrelevant differences between persons.

The criticism then, is as follows. Normative individualism is not a reason to prioritize individual human persons in the context of LP's account. This is because groups are no less *individual persons* than humans are (e.g., Bob is *a* person; the university faculty is *a* person – both are agents that satisfy criteria for personhood). What LP mean by "normative individualism" is actually something like "normative agents-that-matter-ism," which we are

to understand translates to rational human beings.[12] The term "individual" is not doing the work that LP want it to do. The individual versus group distinction should be irrelevant to an account that subscribes to (and heavily depends on) the principle of multiple realizability. LP have both a view that allows groups to be agents (and persons) and a commitment to normative individualism. At best, this combination of views requires much more care in articulating. At worst, these views are incompatible on their account. In sections 5.1.3 and 5.1.4 I will offer some possible pathways to amending or revising LP's account with more explicit taxonomies that capture more fine-grained distinctions.

Only some agents matter and there is no reason to think that all beings that matter are agents. So, if "individual" is meant to qualify the term "agent" or "person," then we still have need of another category for things that matter that are not agents. As mentioned previously, these are traditionally called "moral patients." So, the concept of agency, nor personhood, nor the qualifications of "individual" and "group" are sufficient for capturing morally relevant distinctions that matter, and that, I would argue, our moral categories should capture.

LP have chosen to make use of the "performative" concept of personhood, because it is the capacities an agent possesses and their robustness, dependability, and consistency of performance that makes them persons. They reject an "intrinsicist" notion of personhood, in which what makes one a person is "baked in" to their nature.[13] They can reject that agency is baked in. This is not a problem (infants are born patients but not born agents). But if LP reject that personhood is baked in without decomposing personhood into agency and interests, they miss the relevant baked-in category that overlaps with personhood ("overlaps" in the sense of a Venn diagram, see Figures 5.1 and 5.2). Distinguishing entities that matter, entities with agency, and entities with both, is what motivates the following proposal.

5.1.3 Persons, Agents, Patients (PAP)

"Having agency" and "having morally-commanding interests" are not co-extensive properties. There are some that are not agents, but that have interests that matter. These are

[12] I will go on to lay out categories such that we can account for things that matter that are *not* agents (e.g. infants). Also, in regard to the phrase "translates to rational human beings" - it is understood that the category might be broader, admitting of e.g., angels and Martians, etc.
[13] I am not offering a defense for the given status of human beings having morally commanding interests. I take for granted that the category of beings with intrinsic moral value minimally includes humans.

those who can be morally wronged but are not themselves held to account. They are not agents but are owed obligations from agents. Various terms have been used to refer to this group in the ethics literature, e.g. "moral subject,"[14] "moral recipient,"[15] or "subject of moral concern."[16] I have opted to use the widely used term, "moral patient."[17]

Patients have intrinsic value that entitles them to moral consideration. Agents owe patients this moral consideration; patients are owed moral consideration by agents. Agents have obligations towards patients. Patients are not appropriate subjects of moral praise or blame, as they do not possess the moral duties of agents. Agents are morally responsible for their behavior, but patients are not.[18]

So then, patients are things that matter morally but are not moral agents. We can argue over which things are patients and which are not (e.g., children, chimps, crows, trout).[19] To keep things simple, I suggest holding in mind what I take to be a clear-cut case: human infants.[20]

In chapter 3, we explored the current debate in the philosophy of biology about what it means to be a biological individual. Among the various views explored was a pluralistic approach, in which "individual" can be disambiguated with nuanced categories that better map our terms to the referent (e.g. a Darwinian individual as distinct from an ecological individual). In chapter 4 we justified thinking of biological individuals as agents. Regardless of whether or not we think pluralism about biological individuality has something to offer within the philosophy of biology, I suggest that it may be usefully exported to address the present problem regarding the insufficiency of "agency" when considering what are arguably

[14] Jon Wetlesen, "The moral status of beings who are not persons: A casuistic argument," *Environmental Values* 8, no. 3 (1999): 287-323.
[15] e.g., Vinit Haksar, "Moral agents," In *Routledge Encyclopedia of Philosophy*, (Taylor and Francis, 1998), https://www.rep.routledge.com/articles/thematic/moral-agents/v-1/sections/agents-versus-recipients.
[16] E.g., Fabio Tollon, "The artificial view: Toward a non-anthropocentric account of moral patiency," *Ethics and Information Technology* 23, no. 2 (2021): 147-155.
[17] E.g., Diane Jeske, "Special Obligations," *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), ed., Edward N. Zalta, https://plato.stanford.edu/archives/fall2019/entries/special-obligations.
[18] This leaves the question: do agents have responsibilities towards one another? I have set aside this question. LP says that agents can have reciprocal obligations. We might want to say something like this: agents do not have inherent obligations towards other agents (only patients); but agents *do* have *conditional* obligations toward one another having opted into an agreement. See footnote 31.
[19] For Kant, people are the only subjects of moral worth; there are no non-human agents and patients. Kant, *Groundwork of the Metaphysics of Morals.* See also Robert C. Jones, "Science, sentience, and animal welfare," *Biology & Philosophy* 28, no. 1 (2013): 1-30.
[20] This is not necessarily uncontroversial. See Peter Singer, "Discussing Infanticide," *Journal of Medical Ethics* 39, no. 5 (2013): 260.

importantly different *kinds* of agents (specifically, individual human agents with interests that matter, and group agents, made up of the former, but without interests that matter). In the spirit of philosophers of biology who argue that there is more than one way to count as an individual (i.e. individuals in some sense), in this section I will suggest an important distinction between two different types of agents – agents with importantly different moral statuses. Like the pluralistic approach to biological individuality, the reframing I offer takes the emphasis off of "individual agents" (which would include *any* agent on LP's multiply realizable account, e.g. *group* agents), and allows us to refocus on the categories we care about (but fail to capture when articulated only in terms "individual" versus "group" agency): agents with morally-commanding interests, and agents without morally-commanding interests.

It is important to underline that I am suggesting that pluralism *about biological individuality,* and not simply the idea of pluralism (i.e. about any concept), has something to offer the group agency literature. In both cases (group agency and biological individuality), we are wrestling with what makes something a unified entity - what it is to count as one as opposed to many (and for some, what makes something identifiable and unique). A pluralistic approach offers different *kinds* of unified entities, and thus different ways to count as *an* individual or *an* agent.

So then, by drawing on the pluralistic approach to the debate on biological individuality, we can see a way forward to better match our categories to our projects. Recall the counting problem from chapter 3. How many biological individuals are there in region x? Well, what is an individual? The answer should depend on your purpose: are you measuring fitness or taking an ecological census? Is it only the number of amoebae that matter to your project or is their collective organization and operation as a functional unit also relevant (e.g. slime mold)? Or are you interested only in the products of sexual reproduction (in which physiologically discrete clones are *parts* of an individual)?

The lesson for moral and political philosophy, as inspired by the philosophy of biology, is this: if you want to tally the number of actors in our society – the stakeholders, the contributors to public discourse, the owners and operators of social media profiles - then by all means, include guilds, university faculties, and charitable fraternities. But if you want to

list those in our society who are owed protections and hold the right to life, then do not include among them things like corporations and churches.

It would seem, then, that LP (and others like them) need a more explicit taxonomy which provides categories better suited to distinguishing interests, agency, and coinciding cases (to deal with "the personhood problem" discussed in section 5.1.2). Based on the preceding considerations, there are three categories, two of which need to be distinguished (2 and 3):

1. Beings with morally commanding interests
2. Beings that we hold responsible
3. Beings with morally commanding interests that we hold responsible

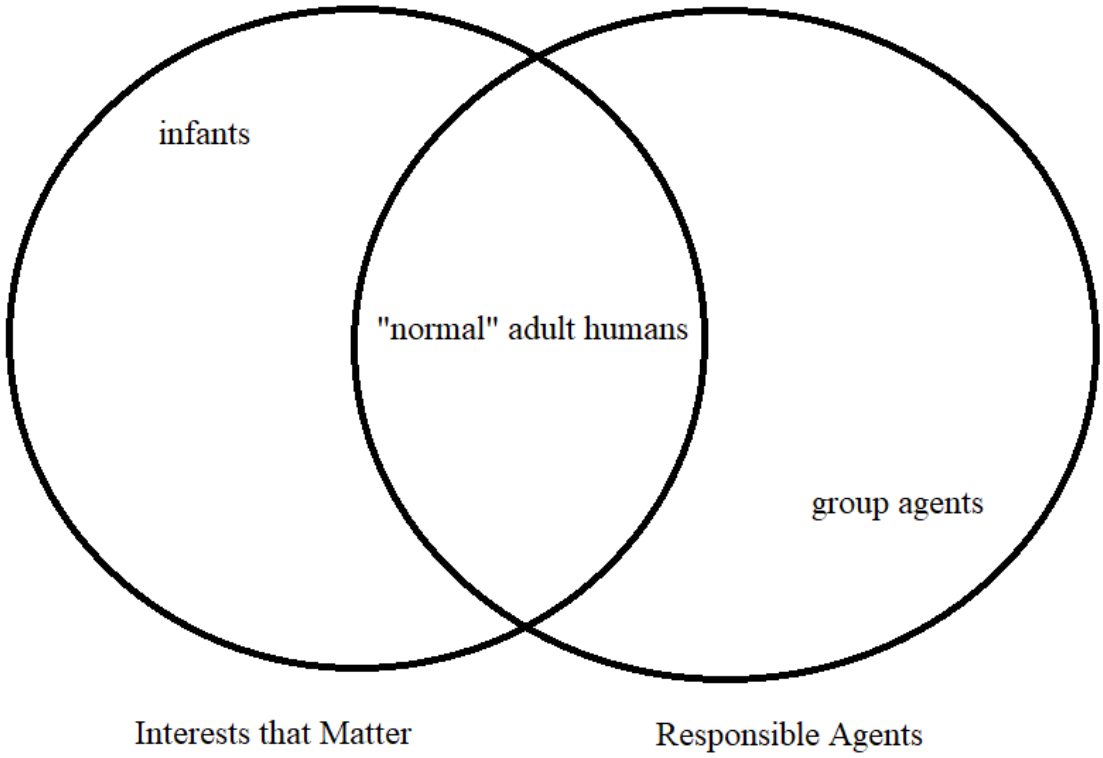

Figure 5.1: A representation of the three categories.

Beings with morally commanding interests that we do not hold responsible are "patients." Categories 2 and 3 are those agents we hold responsible. Yet they are distinct because only category 3 has interests that matter. If we want to keep categories 2 and 3 distinct, we should not label them both simply "agent." For example, either the classical paradigm agent (#3)

should be called agent, and #2 be called something else, or *vice versa* (or alternatively, we could introduce new terms).

I suggest that many will find it natural to apply the term "person" to the traditional paradigm agent (#3). Under this taxonomy, "Persons, Agents, Patients" (PAP),[21] we would have the following:

1. Patients: Beings with morally commanding interests.
2. Agents: Beings that we hold responsible.
3. Persons: Beings with morally commanding interests that we hold responsible.[22]

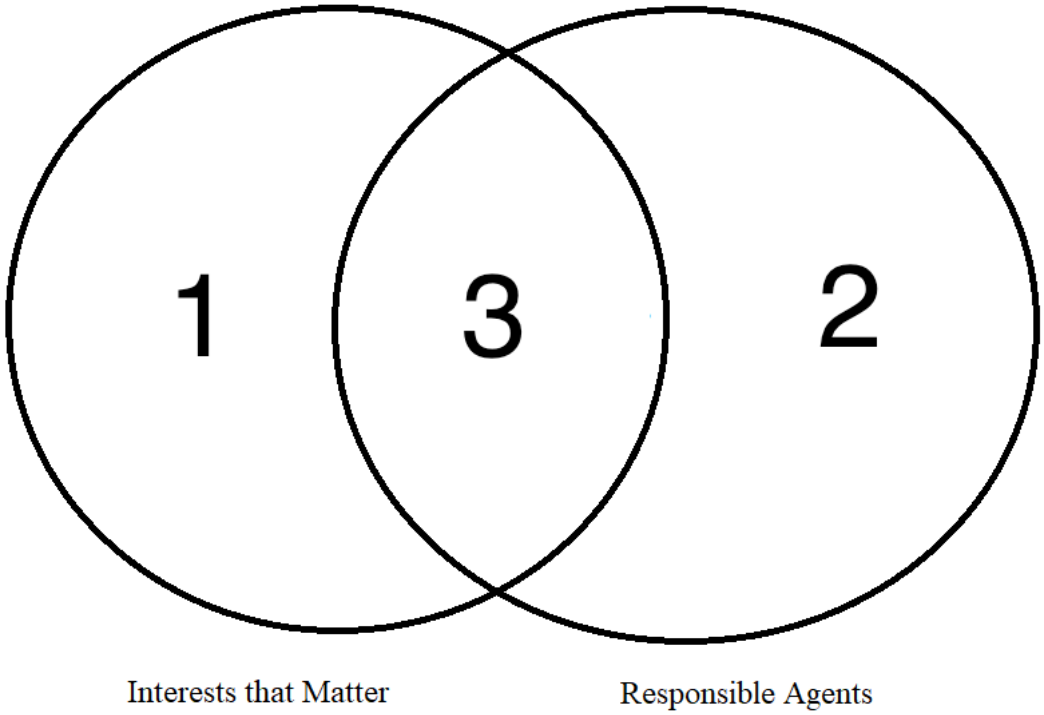

Figure 5.2: A representation of patients (left), persons (middle), and agents (right).

I want to briefly acknowledge, then dismiss, a worry about this terminology. Often in the history of any given discipline, one might be reticent to use words that have historically been used in a variety of ways. To use the word "person" here, even when making clear its use,

[21] Not to be confused with a different "PAP" in the philosophy of action, Franklin's "principle of alternate possibilities." Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66, no. 23 (1969): 829–839.

[22] One may feel troubled that this taxonomy means that, for example, adults with severe cognitive impairments or intellectual disabilities will not be persons. What feels wrong about withholding personhood status from such individuals, is that to do so seems to be saying they don't matter. However, when made clear that these individuals *do* matter because they are *patients*, our discomfort about withholding personhood status dissolves.

risks inviting ambiguity at some point down the line. This sort of worry is amplified in this particular context because unfortunately LP already make use of the term "person," and so, the revision I am proposing cannot simply be offered as a supplement. Contrary to the PAP taxonomy proposed above, LP argue that those without morally commanding interests can be persons. In other words, groups can be persons on LP's account because for LP, being a person just means performing well in the space of obligations. This is most unfortunate because what I think LP want to say about a group's capacities and performances is better expressed in terms of *moral competence* and the capacity to be a *moral actor* (the latter following French, see section 2.1.2). That is, there is some subset of agents that are morally competent and can be relied on to perform as such. Rather than discussing the possibility that some groups which are agents can attain these capacities, LP say that there are some group agents which qualify as group *persons*. So, to return to the concern I am addressing, one might think that, while considered independently that PAP is a good articulation of the three categories that LP's account (and others) is in need of, there is a worry that precisely because PAP was inspired in response to LP's work – i.e. associated with an account that has an alternative definition of "person" – it may contribute to further ambiguity.

However, I would argue that the promise of utility and elegance of the PAP taxonomy outweighs the contingent terminological concerns. I would argue that the clear mapping of actions and interests to the bearers of each is sufficiently clear (e.g. even at a glance, see Figure 5.2). I would prefer to rebrand LP's "personal agents" as "morally competent agents," or simply, "moral agents." This is consistent with the way the account is built up – from minimal agency up through various stages approaching ideal group agents (see sections 1.6 and 2.3). At the least it is neater, as the terms may now uniformly bear the suffix term "agent" (whereas before, there was a clear terminological break with the term "persons"): Minimal Agent → Rational Agent → *Autonomous Agent → *Responsible Agent → Moral Agent.[23]

But neatness is secondary. Most importantly, this change takes what LP call "performative personhood" out of the "person" category and situates it into the "agent" category. Moral

[23] One might read LP and take "*" simply to be descriptions of what comes once robust group rationality is achieved, and therefore not distinctly different types of agents. Note that arrows are used to illustrate the possibility of agential development (i.e. not material implications).

agents are a subset of agents. Persons are a subset of moral agents. Persons are moral agents with interests of moral concern.

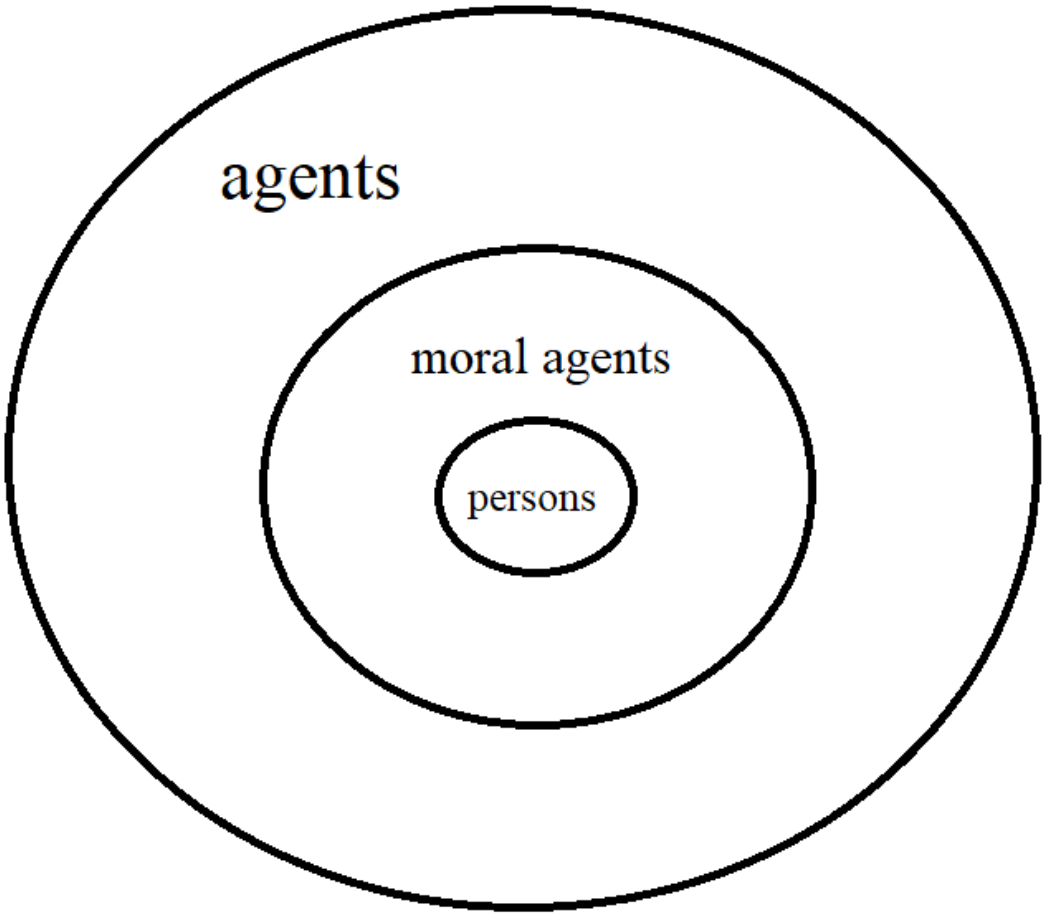

Figure 5.3: A representation of agents, moral agents, and persons.

LP's account is consistent with a world in which there are no agents that matter.[24] The PAP taxonomy makes this scenario easily understandable and tractable. We are imagining a world that has patients and agents, but no persons. If you inhabit this world, you are either a patient or an agent, but not both. This is of course not true in our world. We regularly observe patients that satisfy the criteria of agency (e.g. babies that become rational adults) and are therefore persons. In our world there are clearly beings with interests that matter but are not (yet) agents (e.g. babies). There are also clearly persons that are agents with interests that matter (e.g. adult human beings). The PAP scheme makes it easy and clear then to ask the question: are there beings in the world which are *only* agents? (i.e. non-personal agents)?

[24] Some Sci-Fi stories describe a post-apocalyptic reboot of civilization in which preserved human embryos are grown and reared by non-sentient robot caretakers (e.g. the film *I am Mother* and video game *Horizon Zero Dawn*).

Persons have agency and patients do not. Are there any agents who do not have the intrinsic moral value of patients? This, I would argue, is a better framing *for a project like LP's*.[25]

On LP's view, agency is not sufficient for mattering. Indeed, nor is personhood. In my view, the carving of the three key categories of PAP (person, agent, patient) is in line with the distinct terms offered to capture them. I would argue this is much cleaner than LP's account which would have us separating what matters from what doesn't matter without a tractable vocabulary to do so. On my proposal (i.e., PAP), "person" makes it clear that we are dealing with one who possesses agency and interests that matter – one who can morally wrong another and be morally wronged by another. On LPs account, the term "person" tells you nothing about, e.g. what obligations might be owed to this agent. Is it morally permissible to kill a person? The answer for LP must be, *it depends on the person,* and really, it depends on the *type* of person. The types offered by LP are "individual" and "group."[26] However, again, normative individualism alone cannot do the work here – we must ask: an individual *what*? If individual means "individual person," we have reiterated the ambiguity. LP will not say, "individual human," which gets closer to what they mean (which is "individuals-with-interests-that-matter"), because to do so undermines their reliance on multiple realizability, which is a big part of what makes group agency (and group personhood) plausible.

The key conceptual arguments should not be confused with smaller points about semantic clarity. You could achieve semantic clarity at the cost of an inferior theory. For example, you can say that corporations are persons and that they deserve all the rights that come with personhood (e.g. the right to life, to not be bought and sold, etc.). Alternatively, you could say that "personhood" should be reserved for only those agents that have interests. Both satisfy the *desiderata* of mapping relevantly distinct moral statuses to suitable terminology. But LP are not happy to grant to corporations all the rights and privileges that agents with interests are granted. I would argue that allowing a corporation to be called "person" gives too much credibility to the kind of view LP want to distance themselves from – a view that says, as persons, corporations should receive the same rights and privileges as natural persons (i.e. human agents). I would argue that the word "person" should be reserved for agents that

[25] We should not expect all taxonomies to work for all projects. This is the lesson taken from the work on biological individuality. For LP's project, we need a different taxonomy from the one they have imported from other projects.
[26] Or really, "non-individual," as for LP, "individual" implies a status which e.g., individual robots, may not have.

matter (as proposed in the PAP taxonomy). Of course, this is up for debate, and this kind of argument is not unique to this subject.[27] I would argue that LP offer us the exciting possibility that groups can be persons, but only on a conception of personhood that leaves out that which matters about persons. LP do at one point, anticipate this kind of objection, but remain tone deaf to its import: "we plead guilty to the charge of taking back with one hand what we have given with the other… while we agree that group agents count as persons…. We deny that the rights they enjoy are on a par with the rights of natural persons." [28] The PAP taxonomy can avoid this giving with one hand and taking away with the other.

You might think we have not yet completed the conceptual work with the PAP categories. Suppose I tell you x is an agent. You might ask, which kind? Is x an agent with interests or without interests? Also, I might tell you that y is a patient. You might ask, which kind? A patient with or without agency? What clarifies both of these questions, on LPs account, is to say whether the agent is a personal agent or not and whether the patient is a personal patient or not. However, this is redundant. On the PAP scheme, a more precise term would be offered initially: if an agent were also a patient, we would say "person." If the patient were also an agent, we would likewise say "person." Such ambiguity would not be an issue for the PAP categories, as seen in the disambiguation above in which the categories were simply reiterated as a two-word phrase (e.g. "personal agent"). The PAP categories are useful in avoiding the ambiguity in less precisely descriptive phrases such as "has interests that matter," from which we know not if the referent is a person or a patient. (We can still meaningfully say things like, "only a subset of agents are also patients" or "only some agents are persons," but the referent is the same: person).

The PAP categories are not only precise, but flexible enough to be *usefully less precise*, as when expressing inclusivity. If I am talking about something (e.g. a right) that applies to everyone that matters, then it applies to both patients and persons. Similarly, if I say something (e.g. an obligation) applies to those that are subject to praise and blame, it will apply to agents and persons.

[27] E.g. "We are conscious! We have freewill! But in saying this one must understand 'consciousness' and 'freewill' don't mean what they usually mean..."
[28] List and Pettit, *Group Agency*, 180.

Let us sum up. These distinctions (PAP), if not made explicitly in the group agency literature,[29] are often at work implicitly in and beyond the literature.[30] Adult humans are paradigm agents in philosophy. But that is not *all* adult humans are (i.e. they are agents *with interests*). LP's exciting conclusion that groups can be agents trades on our hard-to-shake intuitive associations of agents with paradigm agents. Yet, what makes agents and persons importantly distinct is simultaneously elided and highlighted in various parts of LP's account (i.e., highlighted in their commitment to normative individualism and elided when arguing that groups can satisfy the same agency criteria as individuals). A commitment to "normative individualism" is needlessly ambiguous when combined with an account of agency in which groups of humans and single humans alike are both *individual agents.* We require a meaningful way to make importantly distinct the two types of agents that matter to us – and further, a category for non-agents with interests that matter (i.e. patients). Within the philosophy of biology, the debate on biological individuality is already well-trodden in this regard. Some physiological individuals are also evolutionary individuals, but others are not. The PAP taxonomy carves up categories that match the distinctions which matter (to LP for their project, and more generally). Specifically, interests and agency are properly distinguished.[31]

5.1.4 Moving Ahead - The Need for Nuance

We have just seen that we can better preserve the distinction between those with and without interests that matter by using the PAP taxonomy. In this section, I will suggest that further

[29] E.g. Killoren, David, and Bekka Williams, "Group agency and overdetermination," *Ethical Theory and Moral Practice* 16, no. 2 (2013): 295-307.
[30] E.g., in legal contexts. See Jon Garthoff, "Decomposing legal personhood," *Journal of Business Ethics* 154, no. 4 (2019): 967-974; John Hasnas, "Should corporations have the right to vote? A paradox in the theory of corporate moral agency," *Journal of business ethics* 150, no. 3 (2018): 657-670.
[31] Based on PAP, agents (aka "mere agents," e.g., group agents) would have no duties towards one another. Agents do not have any interests that matter and therefore are not owed what patients and persons are owed. This might sound counter-intuitive and have implications we don't like. For example, how can we justify saying that agents need to hold to agreements with other agents and refrain from abusing other agents? One response is to say that all of this is a legal, not moral, matter. Perhaps agents do have obligations towards other agents, but only in the legal sense, i.e. our worries are not about morality, but about how we enforce laws within a society (whether those laws are justified is another issue). We could also say something like this: agents *can* have obligations to one another, but not inherently, rather conditionally. Conditional obligations come from entering into an agreement such that no duties existed before the creation of a contract. Other ways of explaining how non-personal agents might owe each other obligations could be cashed out in terms of what consequences violating obligations would have on those that matter (i.e. persons and patients affected by, but not party to, the agreement between agents). But the latter has problems, as the rationale undermines LP's non-redundancy thesis (and similar ideas about non-reducibility in a range of other accounts).

work in this area might be expanded in a similar way, again taking inspiration from the direction things have gone in the biological individuality literature. Drawing on a parallel to the problems of ambiguity and vagueness about biological individuality (see section 3.2), we might think of different kinds of agents in terms of both different types of agency and varying degrees of agency.

We can draw on specific examples from the individuality debate within the philosophy of biology to challenge the idea that agency is the right concept to preserve the distinctions that matter for LP. In biology, we have a plurality of terms to disambiguate the use of the (often rather unhelpful) term, "individual" (e.g. genet versus ramet, replicator versus interactor). What "individual" is pointing to in any given instance is contextual. One is concerned with the "individual" relevant to one's project (e.g. tracking migration patterns or tracking gene flow). Counting the number of hawk cells will not tell you how many hawks are in the area. Including only genetically homogenous entities in one's ecological census may lead one to carve up individuals with very odd boundaries, and as a result, miss the data that matters (e.g. abstracting the "mouse part" from the microbiome, i.e., labelling the two collections of different cell types that comprise the functionally integrated ecological unit of concern). In biology and the philosophy of biology, we have different categorization schemes for different purposes.

In section 5.1.3, I suggested the PAP scheme could feature in a revised version of LP's account. This revised account could properly distinguish those with agency from those that matter, making coherent and consistent LP's motivation and commitment (LP's motivation is to hold some groups responsible for their collective acts and decisions; their commitment is to prioritize and protect persons over agents – "persons" and "agents" understood according to PAP). A highlight of LP's account is that they have made a clear case for an interesting third category not normally, that is, historically, considered in the ethical literature: agents that are responsible for their actions but do not have interests that matter. PAP is a neat breakdown for the well delineated categories of those fit to be held responsible, those with interests that matter, and those fit to be held responsible with interests that matter.

But there is more work to be done. On LP's account, "agency" is asked to do far too much work, picking out a range of entities with significant differences belied by a single title. The simple breakdown as stated above (PAP) denotes agency as being fit to be held responsible.

But even on LP's account some robots and animals will count as simple intentional agents (i.e., minimal agents). As Tollefsen similarly notes, we would not hold *these* agents responsible for their actions.[32] It is natural to ask then, what work the concept of "agency" does if it does not help us make key distinctions? The broad criticism I'm raising, introduced at the beginning of this chapter (section 5.1.1), is about the insufficiency of "agency" to stand in as an adequate description or classification of a wide range of actors. We should reflect: why do we think the concept of agency is a useful category? Is it capable of doing all the work we want it to? A useful concept can be made less useful if unjustifiably stretched to cover all of our concerns. "Agency" in the sense of one that exhibits responsibility fitness is useful for the problems addressed above (regarding LP's concepts of "persons," their commitment to normative individualism, and making distinct beings with morally commanding interests). But what about robots, animals, or loose collectives that only approximate group agents? As this thesis concerns group agency, I will proceed with the latter in mind.

Let us step back just for a second and recap how we got here. LP argue that groups can, under the right conditions, tick all the boxes for agency. Whether implicitly or explicitly, we evaluate candidate agents in comparison with paradigm agents i.e., adult human beings. LP do a good job convincing us that some groups are much more like individuals than we would have thought - groups of individuals can comprise *an* agent. Yet we have already noted that our intuitions may be clouded by the fact that what LP call agency comes together with other very relevant categories based on other features or capacities possessed by paradigm agents - rational human beings - which have interests that matter (and are not merely responsible centers of attitude and action). Add to this the complication that some things will qualify as intentional agents, yet we do not see fit to hold them responsible, which seems to demand the articulation of different kinds or levels of agency (or perhaps more nuanced categories and concepts). Yet recall, LP are trading on the exciting idea that a group can be seen as an agent, similar in relevant ways to individual human agents. One might wonder if this theoretically tantalizing synthesis will, in practice, unhelpfully blur distinctions that matter or e.g., risk ascribing agency to some when doing so makes no difference.[33] In other words, as we go

[32] Tollefsen, *Groups as Agents*, 115.
[33] Sometimes these issues are explored in ethics or philosophy of mind when considering AI, e.g. imagining a number of futuristic scenarios in which non-sentient artificially intelligent humanoid robots are not easily distinguishable from human beings.

beyond PAP to explore cases that call for different kinds of "agents," we will keep the foregoing in mind as a warning not to repeat the same mistakes.

So, while I have heavily emphasized the differences between patients, agents, and persons with the PAP categories, we arguably still have a lot of work to do. The difference between mere agents and persons is a solid case of how eliding a distinction can have severe consequences (e.g. it would be horrific to kill a person we took for a mere agent, or unhelpful and inappropriate to hold a patient responsible for their actions). But there are many less consequential distinctions that matter for conceptual clarity. What do we make of LP's claim about simple robots and animals? Or specific to group agents - how do we describe and categorize the various groups picked out in Stephanie Collins' taxonomy, some of which only approximate group agency (e.g. "teleological collectives," see sections 2.4 and 5.1.5)? [34] That is, while "higher" agents present very real moral problems that we worry about, distinguishing different kinds of agents (and perhaps "almost-agents") across the hierarchy is also an important philosophical problem – a problem that the single term "agent" works to disguise.

The potential for future work here is vast, and again, I suggest that the agency literature could take a leaf from the literature on biological individuality. We need to explore the possibility of new moral and non-moral categories to pick out the loci of agency: minimal, individual, collective, paradigm, marginal, diffuse, intentional, non-intentional, etc. In *what sense* is the focal agent an agent? This is analogous to the question we ask in philosophy of biology: in what sense is a biological individual an individual? Answers may include: in the physiological sense, in the sense of being a unit of selection, in the sense of being the product of sexual reproduction, etc.

We can solicit similarly nuanced senses in regard to agency. Answers to the question *in what sense is the agent an agent?* may include: in the sense of being normatively competent, in the sense of having the capacity to intervene in the environment, in the sense of being rational and autonomous, etc.

[34] Collins, "Collective responsibility gaps," 945.

Incidentally, LP have paved the way for thinking about subcategories of agency. For some types or categories, agency (or whatever we will call it) may admit of degrees, with a "building up" of successive thresholds reached – a kind of agential trajectory (which in some cases may map onto a developmental process of "blooming" or "developing" agents).[35]

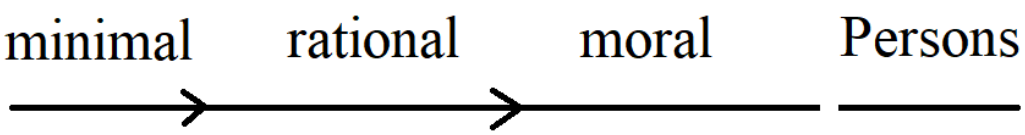

Figure 5.4: A representation of minimal agents, rational agents, moral agents, and persons. Notice the non-continuity with persons meant to indicate a difference in kind, not dependent on capacities achieved or demonstrated, but on intrinsic status (i.e. based on my preferred view consistent with PAP and not LP's sense of "personhood"). The term "agent" is useful here only in the broadest of contexts, i.e., when referencing a focal entity that "does something." The prefixes (e.g. "moral") do all the work in differentiating types of agents.

Each of those depicted above in Figure 5.4 (e.g. minimal, moral) are types of agents (in what sense for LP is unclear).[36] Interestingly, the criteria to attain each type (or "step" in the case of subsequent "development"), may themselves be partially fulfilled at each step. This opens up new possibilities in regard to thinking of "degrees of agency," "degrees of rationality," or "degrees of autonomy" (and thus perhaps even more interpolation with novel "in-between" categories and terms). We can look to philosophy of biology for inspiration about nuanced categories (i.e. taxonomies that capture more appreciable differences) about group agents *qua* collectives and the fact that we can view different groups along a continuum. That is, we are interested in the constitution, composition, and realization of a subset of agents (group agents), and how to distinguish them from non-agents that otherwise seem similar.

I have been speaking about offering nuanced categories and taking a pluralistic approach. Recall the problems of ambiguity and vagueness about biological individuality (in section 3.2). Pluralism is one response to the former; and considering individuality as a matter of degree is one response to the latter. We can generate more nuanced categories to avoid ambiguity, e.g., genetic or ecological individuals. Similarly, we can generate categories to

[35] See section 2.3.
[36] Though perhaps it is clearer on Jensen's presentation. Jensen, "Is our group an agent?"

describe to what extent individuals are individuals, e.g., marginal or paradigm individuals. In both cases, we get a more nuanced classification system better suited to capturing details relevant to our projects and concerns.

Peter Godfrey-Smith's multi-dimensional model might be useful for categorizing different kinds of collectives. What we want are tools to make distinct various criteria and disambiguate senses of the terms we use to describe collectives that meet various subsets of these criteria (examples of what this might look like will be provided in the following subsection). Once armed with these tools, we might then revisit Collins' taxonomy of different types of collectives or Pettit's "three levels of agency" and consider similar (or the same, where applicable) criteria and categories that admit of degrees.[37] Not only might we consider a continuum between marginal and paradigm group agents (following Godfrey-Smith), but perhaps also (or instead), apply these multidimensional models to each of the already nuanced categories available in the group agency literature (e.g. Collins' categories). I will offer examples of these models in the next subsection. Whereas the PAP taxonomy helped make clear the discreteness of important categories (those with interests that matter and those without), what I describe below may help in multi-variable, non-discrete "gray areas" between continuous categories. Again, recall the problems of vagueness and ambiguity from section 3.2. Just as considering the problem of ambiguity about biological individuality led us to the PAP taxonomy; considering the problem of vagueness about biological individuality inspires us to consider agency (and the criteria constitutive of agency) as a matter of degree. Both give rise to new, more precise categories.

Measuring degrees of agency (or individual capacities thought to be constitutive of agency) might help address the ambiguity of inferring an "inexplicit" decision procedure or in our investigations of possible non-intentionally formed group agents (the latter a terrain largely unexplored by LP).[38] While explicit decision procedures are evident and identifiable by definition (e.g. a formal vote), it is less clear whether some groups operate with an implicit decision procedure (e.g. something sufficiently like an explicit procedure to justify attribution of agency or a degree thereof). We might worry that a permissive account of "inexplicit" will

[37] Pettit, Philip. "Corporate Agency," 249.
[38] E.g., if an individual is not aware that they are in a group agent, they might also be unaware that they are partaking and agreeing to a decision procedure, e.g. 5th generations groups, repeating the processes of their ancestors without certain knowledge or intention among individuals.

over generate agents, e.g. from eye contact and head nods among neighbors or work colleagues.[39] Other times, we might over generate responsibility gaps – not holding a collective to account when perhaps we should have. For example, perhaps Pettit's "single action" scenarios (e.g. people at the beach rescuing a drowning swimmer), or Collins' "teleological collective" (e.g., a grass-roots protest group) could be plausibly held more or less responsible for collective actions to the extent that they exhibit capacities plausibly qualifying for a position closer to a clear case of an inexplicit decision procedure, as opposed to a minimal or marginal case in which a decision procedure could not be justifiably inferred.

In the special cases of non-intentional group formation and inexplicit decision procedures, we may also want to deviate from LP's strict criteria. This is because while a decision procedure is definitional of a group agent, to the extent that we have more nuanced needs than the classification of "agent" or "not agent" can satisfy, we may also choose to weigh the primacy of a decision procedure with other factors, e.g. if all individuals agree on a belief, a goal, and jointly interact based on a belief in that goal. If their degree of "togetherness" (or whatever we might call it) is high in this regard, and is coupled with what plausibly, but not certainly, looks like an inexplicit decision procedure, then we might be willing to attribute responsibility to the collective based on their overall "responsibility fitness score" (and can imagine doing so without mention of the term "agent," e.g. "agency score").

Finally, I will just take a moment to state explicitly what we have been circling around in the last few paragraphs. What might motivate this quest for nuance is one of the biggest concerns in the wider group agency literature: that of responsibility gaps.[40] Recall that responsibility gaps are more generally tied to collective action problems: how do we apportion responsibility to the consequences of a collective act when all the individual actors made nominal contributions? While LP have provided a framework in which rational groups with decision procedures can be held responsible, there remain plenty of situations in which we might want to be able to redress less organized groups. There may be cases in which, after holding key individuals responsible, there remains a surplus of unattached blame – situations in which the scale of harm brought about by collective action is not satisfactorily pinned on a

[39] The kind of scenarios described by Pfeiffer in response to French (section 2.1.2). Pfeiffer, "The central distinction in the theory of corporate moral personhood."
[40] This underlying concern motivates much of the literature. However, the terminology differs, e.g. "responsibility deficits" in Hasnas, "The Phantom Menace of the Responsibility Deficit."

couple of key organizing actors (e.g. protest organizers), or cases in which *no* single individual who contributed to the harm did so with enough impact to hold individually responsible. Perhaps this problem cannot be satisfactorily addressed from within LP's account as it stands (in which avoidance of responsibility gaps is limited to cases featuring groups that satisfy LP's rigid criteria). If we are intent on further wading into these muddy waters, we may need to think about our categories and their criteria a bit differently. In doing so, we might find a suitable foundation on which to launch a more nuanced account of relevantly distinct groups. Or perhaps not. My modest suggestion offers one way to move the exploration forward.

5.1.5 Representing Agency and Agency Criteria as Admitting of Degrees using Multi-Dimensional Models

In the previous subsection (5.1.4), I suggested some ways that ideas from the literature on biological individuality might help us address the need for more nuance within the group agency literature. Let us take a closer look at how at least some of these ideas may be put to use. First, I will consider a couple of examples based on criteria offered by LP, then look toward a wider application making use of Collins' categories.

Below are a couple of examples of how fulfillment of LP's criteria for rationality and responsibility could admit of degrees, allowing for a more nuanced categorization of various groups (and perhaps candidate agents more broadly).[41]

[41] For a reminder of the role these criteria play in LP's account, see sections 1.2 and 1.7.

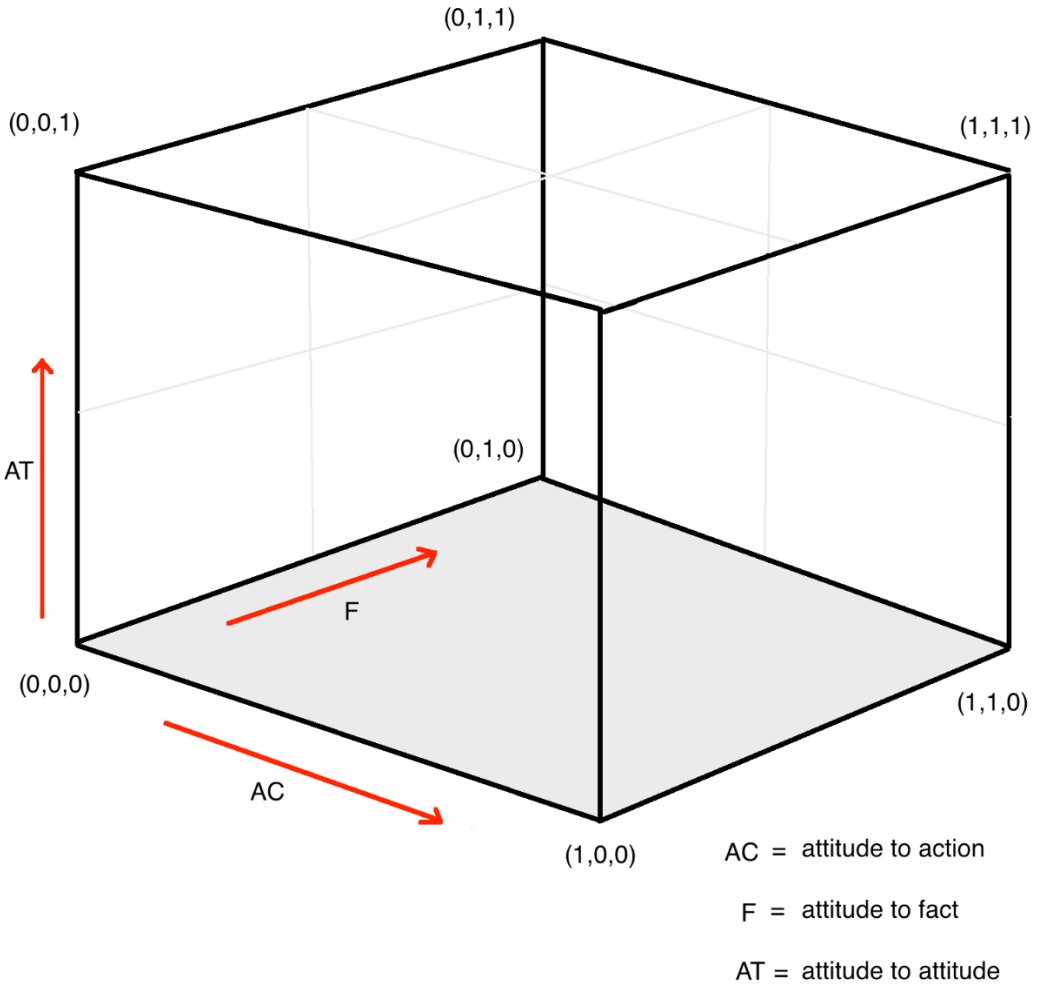

Figure 5.5: A model of rationality. Figures 5.5 and 5.6 show three-dimensional models of rationality and fitness to be held responsible. If we allow that rationality or responsibility fitness could *in principle* admit of degrees, these models may be of use in our attempts to meaningfully categorize more or less rational or responsible groups.

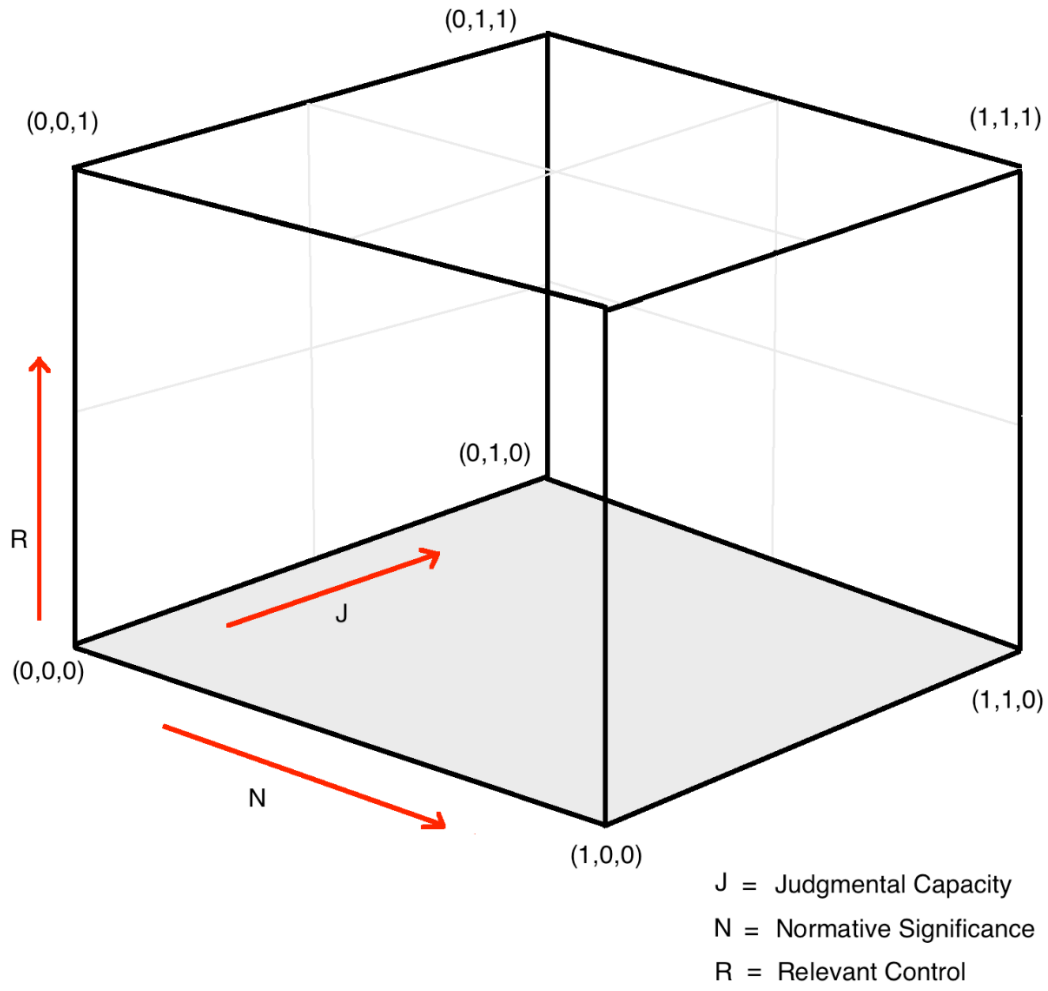

Figure 5.6: A model of responsibility fitness.

These three-dimensional models are inspired by Godfrey-Smith's measures of biological individuality (also used as measures of Darwinian populations). Based on the placement within the cube, we can infer the degree of fitness to be held responsible or degree of rationality, e.g. marginally rational, paradigmatically rational, or somewhere in between (paradigm cases located in the upper right corner, at coordinates "1,1,1"). Of course, considering rationality and fitness to be held responsible as other than binary may already be controversial for some.[42] But on the other hand, these types of models may be a useful resource for messy situations involving various collectives for which "agent" and "not agent" are not sufficiently helpful.

[42] However, note that it isn't a problem for individual elements represented on individual axes to be binary (1 or 0, e.g. "normative significance").

To develop this a bit further and give some examples, consider the rationality cube (Figure 5.5). There may be some groups that exhibit means-ends consistency, represent facts about the world accurately, but hold inconsistent beliefs (e.g. a budgetary committee that collectively agrees on a budget, but for different, incompatible individually-held reasons) - occupying coordinates "1, 1, 0". There may be other groups that act with coherent practical rationality, hold a consistent cannon of internally coherent beliefs, but are wildly at odds with foundational truths about the world (e.g. a new age cult acting with solidarity on nonsense[43]) - occupying coordinates "1,0,1." Others may be theoretically consistent but fail to coordinate properly such that actions follow from their beliefs and goals (e.g. a hobby club good at generating opinions, but consistently failing to align group decisions and actions with their views such that the former follow from the latter) - occupying coordinates "0, 1, 1".

Figures 5.5 and 5.6 are inspired by LP's account, drawing on their criteria for rationality and responsibility. Such models might be useful for those wanting to amend or build on LP's account. However, multi-dimensional models may be put to use more broadly. Therefore, it is worth thinking of at least one example of how multi-dimensional models might be applied beyond LP's work. The model below in figure 5.7 is one way in which we might account for groups that possess aspects of agency such that they *approximate* being agents.

[43] A religious group that launches a coordinated campaign to tirelessly transfer members' spiritual energy into crystals for storage and retrieval by aliens.

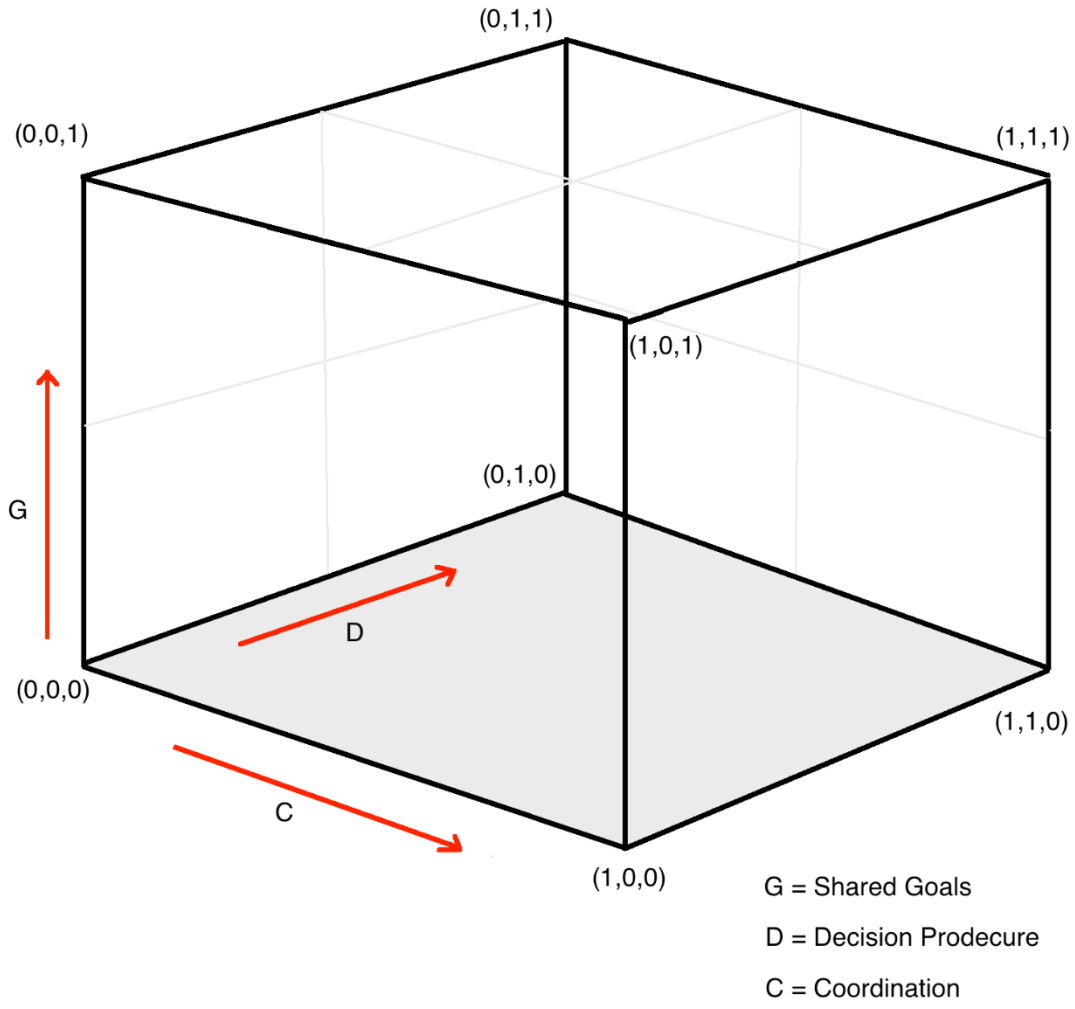

Figure 5.7: A model of collective agency.

We can represent within the cube depicted in Figure 5.7 whatever novel terms and categories we want to capture various kinds of groups. For example, we could use Collins' categories presented in section 2.4, distinguishing groups based on whether or to what extent they share goals (G), coordinate their efforts (C), and operate with a decision procedure (D). Diffuse collectives (e.g. gamblers) would be located at coordinates "0,0,0." While the gamblers all want to get rich, their goals are "shared" only in a weak sense – they are not realized by a single state of affairs (see section 2.1.1).[44] Teleological collectives (e.g. spontaneous beached-whale rescuers) would be located at coordinates "1,0,1." The rescuers share a goal and are responsive to each other's actions but lack a decision procedure. Well-functioning corporations would occupy the paradigm slot at coordinates "1,1,1." Of course, some axes

[44] Tollefsen, *Groups as Agents*, 29.

may be best understood as ideals – perhaps no group ever reaches "1" as the perfect fulfilment of some criterion. This aspect of the model should be thought of as a feature, not a bug. Realistically, groups will be "located" like pockets of gas or clouds within the cube (i.e., a portion of the cube's volume, not a point) and not perfectly in any corner or along any side. In such a case, a threshold for the satisfaction of whatever criterion an axis measures would be set at a value below the ideal (i.e. < 1.0).

A model like the one above (Figure 5.7) may be consulted together with a model of fitness for responsibility (whether the LP version in Figure 5.6, or another). It must be stressed that we are in very messy territory (plagued by conceptual vagueness and uncertainty). However, these clumsy tools may be all we have for messy situations. Consider the January 6th (2021) storming of the U.S. Capitol (recent at the time of this writing). From what I have thus far gathered from the news, no one has seriously suggested holding *the mob* accountable. More sensible efforts have been directed towards identifying and prosecuting certain individuals identified on videos from the event, as well as the president at the time (on the grounds that he played an instigating role). And yet, suppose congress makes a compelling case for the historical importance of formally censuring the mob, and doing so in a non-distributive way (even without the theoretical justification we would normally require to treat the mob as a meaningful unity - some protestors were peacefully demonstrating outside the Capitol while others breached the building). Suppose that based on the details, we could locate "stormers of the Capitol" at the coordinates "1, 0, 0.5" in Figure 5.7 (i.e. something like, "a mixed-goal, coordinated collective").[45] Again, while messy, for purposes such as these, we might say that while possessing nowhere near paradigm group agency, the stormers of the Capitol reached a sufficient threshold of approximating agency to justify and render meaningful censuring them as a group. More severe censure or punishment could not be justified, but, we surmise, "1, 0, 0.5" degrees of agency is enough to warrant a censure which, according to reasons x, y, and z, is historically important to go through with. I am not committed to this particular historical example, but rather just attempting to illustrate, in principle, how "agency scores" might be useful for thinking about agency as Godfrey-Smith argues that analogous scores are useful for thinking about biological individuality.

[45] These could alternatively be expressed within a range in consideration of both empirical uncertainty and conceptual uncertainty (e.g. was an inexplicit decision procedure in play warranting a score of "1. 0.5, 0.5"?).

While space does not permit a full exploration here, the above example allows us to ask about a range of interesting cases in which agents contain other agents. We have seen how an account of agency might not permit nested agency (e.g. Rovane's, see section 2.2.2): one is either part of an agent or an agent, but not both. When considering collectives that exhibit agency (or some aspects of agency) only to a degree, conditions of nestedness are then up for debate. That is, we cannot say with certainty whether or not an agent nests within another agent if either's agency is up for debate. Assuming an account that does not rule out such cases, we could explore situations in which there are group agents within looser groups approximating agents (e.g. clubs and fraternities within the mob). Would higher or lesser agency scores for some collective affect the agency score of another group in which it nests (or contains)? If so, it would be interesting to compare and contrast such cases with cases of biological individuality on a multi-level selection account (e.g. Clarke's), in which one might be more or less an evolutionary individual to the extent that individuals in which one nests (or contains) are themselves more or less individuals. Or in consideration of what has been laid out previously in this chapter, different kinds of agents or individuals may contain others: physiological individuals may contain evolutionary individuals (e.g. holobionts containing bacteria, i.e. two or more evolutionary individuals that make up one physiological individual) just as group agents may contain persons (agents that matter within agents that don't – the former making up the latter). There are a range of interesting cases of nested agency and individuality worthy of future work on the topic (see section 6.2.5).

5.1.6 Summary and Conclusion

In this section, we have explored how the topic of how biological individuality might meaningfully inspire and inform theories of group agency. The general lesson we have learned is that the concept of agency is insufficient to preserve what matters in some contexts. Inspired by pluralism about biological individuality, we can take care to avoid coarse-graining meaningful categories, for example those with different moral implications. A more specific lesson is offered in models of how we might capture these nuanced categories with "scores of agency" or similar ratings for each criterion of agency (or a kind of agency, etc.). Thus, reflecting on the *problems* of biological individuality (e.g. vagueness and ambiguity), has inspired possible *solutions* (or plausible inroads) to the ongoing work on group agency.

5.2 Group Agency Informing Biological Individuality

Let us now move in the opposite direction. Just as the group agency literature can benefit from the work on individuality within the philosophy of biology, so might those working on biological individuality benefit from the work on group agency within moral and political philosophy (and other areas that deal with agency, e.g. philosophy of mind).

We are now familiar with thinking of both agency and biological individuality as an achievement of unity (sections 2.2, 2.3, 3.3.1, 3.4.2, and 4.4.1). Group agency is forged by individuals who organize their activities to maintain rational unity at the group level. I will argue that considering some of the incentive-compatibility problems that arise in the context of group agency may shed some light on how to better approach biological individuality. Some ways of conceptualizing and reasoning about group agents allow us to see certain threats to group level unity as explanatorily relevant (e.g. on LP's account). When we apply this group agency perspective to biological individuals, we are better equipped to give fuller explanations – explanations that we are denied when simply offering criteria for individuation, and then enumerating individuals (even of different kinds). The descriptive and explanatory tools of agency are available to us even when we are unsure of what technically counts as a biological individual of some kind, at some time.

5.2.1 The Insufficiency of Individuality

For many interesting cases, we will not limit the application of our as-if agency heuristic to individuals; we will instead (or also) think and speak about the entities that make or break those purported individuals (i.e. those entities that contribute to the unity or degeneration of individuals). Indeed, it is because individuality is what is at stake in these cases, that agency concepts are useful for investigations of biological individuality.

As we saw in chapter 3, many philosophers debate about how to best understand and characterize biological individuality. Explanations of what makes an individual are often explored like a blueprint or recipe – when some collection of living matter ticks all the boxes, it counts as an individual. This static picture does not tell the full story. Many of the cases that interest philosophers of biology are cases that feature different kinds of individuals at different levels of the biological hierarchy (e.g. organisms and genes), entities that exhibit individuality to a degree (e.g. somewhere between a collection and an organism), or apparent

individuals with heterogenous components (e.g. multi-species consortia). We sometimes observe entities that puzzlingly appear at odds with themselves – a disunity that calls into question the unity of the entity at the focal level.

Recall that one fundamental requirement of *agency* is the sameness of individuals over time (see section 2.2). The concept of agency may be better suited for fuller biological explanations in which focal traits and behaviors are mapped to the relevant agents and their goals – mappings that may surprisingly span hierarchical levels, timespans, and series of interactions.

Group agency *in particular* offers a multilevel perspective where we intuitively treat the parts of agents as agents themselves, and where this recognition significantly factors into our descriptions and explanations. If we focus on individuality alone, then we may fail to capture the relevant fact that individuals contain or are contained within other individuals. This nestedness is not merely a relation of composition but of potential interaction (i.e. up and down). The observed behavior of any individual does not necessarily invite an explanation given in terms of *that* individual. Add to this, that these nested biological complexes have relations to other such complexes at multiple levels, over time and space (i.e. side to side, forwards and backwards, etc.).[46] As I will argue, being precise with individuality concepts is a great start, but their deployment may be under-explanatory for the aspects of some individuality cases that interest us most. Fuller, more satisfying explanations will include descriptions of agendas and goals – and importantly, not only the agendas and goals of the focal, purported individual. I will mostly focus on particularly interesting cases of disunity (or apparent disunity). I aim to show that the explanations for a behavior or trait we observe of an individual at one level, will in some cases, feature individuals at another level. Specifically, and in terms of agency, the *behavior* of an agent at one level may be understood as the *act* of an agent at a different level. Further, it is often the relations between individuals, described in terms of agency, that contextualize the "individuality" of each. So, while we may heuristically think of individuals as having their own interests and goals (section 4.3), we will sometimes want to explain individuality (or lack thereof) in terms of agency *within* an individual (or agency of a higher-level entity of which the purported individual is a part).

[46] I'm saying "complexes" to acknowledge the potential multiplicity of entities that might be constitutive of individuality, while avoiding the specification of which chunks of matter count as "agents" or "individuals." In section 5.2.5, I will say more about the phrases, "forward and backward," "up and down," and "side to side."

Let us begin by exploring two cases, which in light of the material covered in chapters 3 and 4, we can appreciate as relevant to the topic of this thesis. The first case will be new; the second was introduced previously in chapter 4.

Case One: How should we understand cancer?

In a multicellular organism, each cell's evolutionary interests are aligned - they are genetically identical.[47] While a population of multicellular organisms (e.g. horses) can participate in natural selection, the population of particular cells that make up each multicellular organism (e.g. horse cells) do not. The cells that make up a multicellular organism are not in evolutionary competition – they do not make up a Darwinian population. The cells share, as an aggregate, a unity of purpose. Or so we normally assume.

Random mutations (e.g. DNA transcription errors) result in cells that are no longer genetically identical to others in the multicellular organism (by definition). This results in the potential for "rogue cells" multiplying according to their own evolutionary interests, at the expense (and potential demise) of the multicellular organism (and so ultimately their own).[48] Competing cell lineages participate in natural selection at the cellular level, *within* the multicellular organism. The population of cells, now with distinctive variants, has become "visible" to natural selection. In such a case we have two (or more) individuals at cross purposes with one individual nested within the other – physical continuity without purposeful unity.[49] Pradeu writes, "cancer is a disease that results from a dysfunction of the mechanisms that normally insure the cohesion of the multicellular individual."[50]

Case Two: How can we explain the frequent appearance of male sterility in flowering plants?[51]

[47] Bracketing, for simplicity, microbial symbionts, transplanted organs, mosaicism, etc.
[48] With the rare exception of transmissible cancers.
[49] At this point in the thesis, I will be freely referring to goals and purposes without the scare quotes. For a reminder on why this is useful, justified, and not problematically teleological, see chapter 4.
[50] Thomas Pradeu, *The Philosophy of Immunology* (Cambridge University Press, 2020), 30.
[51] Observed in over 150 species. Patrick Schnable and Roger Wise, "The molecular basis of cytoplasmic male sterility and fertility restoration," *Trends in plant science* 3, no. 5 (1998): 175-180. See also, Steven Frank, "The evolutionary dynamics of cytoplasmic male sterility," *The American Naturalist* 133, no. 3 (1989): 345-376.

Recall this example from chapter 4. The trait of sterility has a function – but this function does not benefit the plant (rather it benefits a component of the plant, the mitochondria). It makes little sense to think of sterility as a means to fitness. Like the case of cancer above, we seem to be observing instances of "dissociated" organisms.

We might be satisfied that a horse and a Plantago plant are both individuals if not for the confounding cases presented above in which these individuals seem to be at odds with themselves. These cases undermine our attempts to clearly apply our agential heuristic at the focal level (to the multicellular organism in case one, and to the plant in case two). Indeed, we do not normally stop considering them individuals (i.e. a horse is still a horse – we just say *a horse with cancer*). In response to such cases, some have proposed counting different numbers of different kinds of individuals, e.g. the horse is two genetic individuals but one physiological individual.[52] However, we might suspect that "individuality" is not enough. How can we *best* talk about these cases (and others we find similarly perplexing)? It would seem our explanations need to include details about something *happening;* and what is happening involves the relevant entities doing things – interacting, competing, or cooperating. It is not satisfying (e.g. in the case of cancer) to say, at time t1 there was one physiological individual and one evolutionary individual, but at time t2, there was one physiological individual and one evolutionary individuals. If our stories are to help us to make sense of what is happening, then we need continuous characters who persist through puzzling cases (as opposed to individuals popping in and out of existence). *As-if agency* allows us to tell such a story without making any particular ontological commitments.[53]

Often biological individuality is debated in the abstract realm of criteria and categories, removed from the relational context, of interactions among intra-individual components or between individuals, that forges the entities we would call individuals (of whatever kind). Agential thinking allows us to see the different goals and interests at play *within* an individual (or purported individual) over time, and group agency thinking provides a framework to do this at multiple levels – when agential heuristics might be applied to both parts and wholes. I aim to show that agency concepts are working implicitly in our understandings of

[52] E.g., Wilson, DiFrisco, Godfrey-Smith.
[53] What matters is the continuity of the agent through actual events, not their ontological status.

individuality already. I would argue that portions of the biological individuality debate might best be clarified by making our agential thinking explicit.

In cases like those above, we want to identify the point where applying our agential heuristic to a single level (e.g. the horse, the plant) went wrong (i.e. in which the agent we picked out lacks the unity of *an agent*). We need something more than what carving out different individuals alone can give us. Our investigations risk missing the proper emphasis when taking up a troubleshooting perspective (in which we look to the components of an individual like broken parts). In Dennett's terms, doing so means taking a "design stance" or "physical stance" towards the individual rather than an intentional stance. The difference is this: when investigating a stopped clock, we expect to find a broken or misplaced piece, but we are not guided in our investigation by a suspicion that some spring has defected in fulfilment of its own goals at odds with the clock's.

Rather than abandon an intentional stance at the troublesome focal level, we could alternatively maintain an agential perspective (i.e. remain in the intentional stance) but just apply it to a different level of the biological hierarchy, as well as taking up a diachronic, rather than merely synchronic, perspective (this view will be developed in the subsequent subsections). I will argue that what may aid philosophers of biology to take up this approach more consistently is a kind of reasoning that comes naturally when examining group agents.

5.2.2 Agency

I have just suggested that identifying and characterizing biological individuals might often go best when we adopt a multi-level agential perspective. This more particular agential perspective presumes the utility of *any* agential perspective used in a biological context. Recall from chapter 4 what was argued about agential thinking more broadly (i.e. generally, not specifically about group agency): to conceptualize some biological entities as rationally unified actors with purposes pursuing goals *via* various strategies is useful and justified. Specifically, agential thinking supplements a fuller picture than the debate on biological individuality alone can offer. Talk of even a plurality of biological individuals does not capture what we care about in some cases – for these cases, agential thinking is required. The material in chapter 4 served the purpose of building an argument – demonstrating agency and unity can justifiably be taken as common currency between these two bodies of literature.

However, because thinking of, e.g., organisms, as rational agents has in recent history been shunned by biologists who worry that anthropomorphizing the biological world will bring with it the historically associated views of directedness, backward causation, or theistic purposiveness, the acceptable use of agential talk is itself an important lesson learned for those currently working in biology and philosophy of biology. So then, philosophers thinking about biological individuality can benefit from how philosophers think about agency.

More specifically, philosophers of biology working on the problem of biological individuality can benefit from moral and political philosopher's theories about *group* agency. What the group agency literature in particular provides is a context in which, like biological individuals, agents may exist in nested hierarchies (and if not, the reasons why are relevant). Agency concepts are nothing new in biology – they have been used and misused for centuries. However, the notion of biological individuality as the evolved character of a collection of lower level individuals is a more modern insight.[54] If imported into the philosophy of biology, the concept of group agency would bring with it the working assumption that the parts of a group agent should be usefully thought of in terms of their own interests and goals. The unity and robustness of group level agency is suspect from the start, whereas biological individuality is taken as given (e.g., the theory of natural selection takes for granted the existence of biological entities that, e.g. vary, reproduce, compete for resources and pass down traits). Applying "group-agential thinking" to biological individuality highlights the way we think differently about biological individuals and the parts which make them up (e.g. multicellular organisms versus the cells that make them up).

So then, how is it that we think differently about biological individuals and their parts versus group agents and their parts? Our natural bias when considering group agents is that the parts are presumed agents, while the whole may or may not exhibit its own agency. This bias might provide a balancing set of considerations when thinking about biological individuality. With group agency, we default to thinking of individuals as the *real* agents, while skeptically considering whether the group of individuals is also an agent (or whether individuals *contain* agents). With biological individuality, we take it as given that the focal entity (e.g. an organism) is the real individual, often bracketing a similar stance we might also have of its parts (or a group of individuals). As I aim to demonstrate throughout this section, multi-level

[54] Maynard Smith and Szathmary, *The Major Transitions in Evolution*; Buss, *The Evolution of Individuality*.

agency helps in our investigations and explanations, on an up to date understanding and more in-depth appreciation of biological individuality. The intuitions brought to both debates, biological individuality and group agency, can support one another, providing different ways of thinking about relations between hierarchical levels.

In the previous section on group agency (5.1), we discovered that we were especially concerned with one kind of individual. Indeed, limiting our focus to agents missed the distinction between agents that matter and agents that do not – and the term "individual agent" was no help.

For our considerations of biological individuality, the opposite is true. Often it will be most relevant, not what kinds of individuals we could (in principle) distinguish, but rather what individuality-constituting (or individuality-degenerating) agents are actually in play. Within the context of group agency, people are presumed agents while group agency is a remarkable innovation and achievement. In the context of biological individuality, it is the remarkableness of intra-individual achievement we risk overlooking (i.e. the unity among the parts is easily taken for granted).

In the case of group agents, unity and disunity are always explained in terms of interagency. Individual people succeed or fail to agree, align their incentives, and sufficiently coordinate their deliberations and activities. Yet in the case of a biological organism we are much more likely to describe unity in terms of evolved traits and functions and disunity in terms of dysfunction. The case of male plant sterility shows how this leaves out the most interesting part of the story: a trait's function should be indexed to the proper individual. Our reasoning and articulation of doing just that takes the form of an agential explanation: it is the tale of two intraorganismal agents at cross-purposes, exhibiting conflicting traits as strategies towards different goals. Hence, the disunity of a stamen with no pollen. The failure of the plant's unity was, in the case of male sterility, a success for the mitochondria.

5.2.3 Shifting of the Intentional Stance

Recall why LP argue for the reality of group agents. The answer is a specific version of a much more general point about our rationale for considering something to be real. The rationale can be easily illustrated. When relating to a friend that you recently attended a

departmental philosophy seminar you do not say, “while inhabiting a familiar environment I observed a novel low-entropy homeostatic chemical system repeatedly propelling puffs of air towards the audience.” This may be true but does not capture the relevant information. You need to employ higher-level descriptions to capture the information that matters and relate it to your friend. Concepts such as “building,” “room,” “chair,” “guest speaker,” “presentation,” and “projector” are justified by virtue of their indispensability to our predictions and explanations. We risk underreporting or misreporting things and events in the world by not acknowledging the reality of these higher-level entities. Describing the evening’s events at the level of elementary particles does not help your friend understand why the speaker’s reply to Sally’s question during the Q&A provoked irritability in Mark, whose recalcitrant skepticism led the speaker to go ten minutes over the allotted time.

The contribution of accounts like LP’s is that they enable us to say that there is a real sense in which a corporation is an actor that makes decisions, enters into agreements with others, deliberates, and pursues ends. To describe corporate actions at the level of the individuals that make up the corporation is to make the same kind of mistake illustrated above (for reasons made clear in Chapter 1). Talk of group (or higher-level) agency need not be dismissed as merely metaphorical.

Just as we would miss the significance of some events at a higher level if we didn’t recognize the agency of groups, so might we miss the right explanations if we incorrectly attribute to groups what should be attributed to individuals. In the following section I will argue that some of the problems of LP’s group agency account, might offer a solution to problems about biological individuality.

Problems for Group Agency – Disunity

Threats to group level unity are such that they may present problems for accounts of group agency. I propose that the problems that plague group agency accounts (LP’s in particular), particularly in regard to the alignment of interests, will inform the useful agential heuristics we apply to the biological world. So interestingly in this instance, it is not the successes of the group agency literature that hold potential utility for the biological individuality debate, but rather the failures.

Here I will propose some new potential problems for LP's account of group agency. Then I will suggest what the examination of these problems has to offer philosophers of biology when considering biological individuality.

"Agent when" Groups

We can both imagine and find real examples of instances in which individuals create a business venture simply to profit from its ultimate failure and dissolution. A paradigm case might be "pump and dump" stocks, or more recently, similar schemes involving crypto currencies.[55] Cryptocurrency investors create and spread hype about the next big thing prior to an "initial coin offering" (ICO), buy while the price is low following the ICO, and after others hop on the bandwagon and the price increases, they sell everything, "bursting the bubble" having withdrawn their investment and leaving the other investors they enticed to choke on the dust.

Or consider hypothetically a short-lived firm that does shoddy work with cheap materials. They finish the job then dissolve the company. The same day the company is dissolved, a new one is created with the same employees and assets (conveniently purchased from the prior company's liquidation). Acme Construction dies and ABC Construction Company is born. When your house falls down around your ears it will not be ABC's fault – after all, it was Acme that was contracted to build your house. Acme no longer exists. These sneaky individuals profit from a joint strategy to create and kill off the legal entities they briefly comprise.

With these examples in mind, we can imagine a loose coalition of individuals that are specifically exploiting the criteria regarding what makes and does not make a business a single (continuing) entity. That is, these individuals have made group agency itself the true focus of their venture, propping up their endeavor variously as the undertaking of a single agent, or the undertaking of a collection of individuals, as it suits them. Call such a group an "agent-fluid meta group agent,"[56] or more simply an "agent-when group."

How would this work? By changing their decision procedure from premise-based to conclusion-based (recall that roughly, the former ensures group level rationality required for

[55] J.T. Hamrick et al., "An examination of the cryptocurrency pump-and-dump ecosystem," *Information Processing & Management* 58, no. 4 (2021): 102506.
[56] "Meta" because the group agent has, as its subject and orientating focus, the criteria for constituting a group agent (i.e. the group agent is *about* group agency) – fluid because the group "dips" in and out of agency.

agency while the latter does not), this group of individuals (i.e. an "agent when group") could change their status from group agent to non-agential collective. That is, they could carefully engineer which decisions belonged to the group agent by virtue of engaging in a decision procedure that ensured robust group level rationality; or, if more beneficial, could come to take group action as a collection of individuals. With a change in structures and procedures, the agent-when group can flick on and off an "agent switch." This way they could reap benefits that groups, but not individuals alone, could achieve – privileges that could be distributed to the individuals in the group. More worryingly, they could avoid responsibility as individuals by making some decisions as an irreducible group (using the premise-based decision procedure). Then perhaps at other times, they can avoid group responsibility by making other decisions as individuals (using the conclusion-based decision procedure). They can manipulatively mix and match, exploiting the "on/off" nature of their endeavor. When the group dissolves, they escape liability as individuals – for as LP demonstrate, we cannot, without additional information, infer any one individual's contributions from group level outputs.

The problem being outlined here is of course more severe than those that can be sorted out legally (in which you could say the *real* group agent was the long-term beneficiary of propping up and dissolving various companies). The problem with agent-when groups concerns the inability to attribute *moral* responsibility. Agent-when groups are not simply playing with titles or facades (as in a legal context with the construction example). Agent-when groups are more pernicious; they are literally coming in and out of existence as agents according to LP's criteria.

An additional way for "agent-when" groups to carefully reap benefits and avoid responsibility is by not only varying decision procedures to fluctuate the group's status (unified agent versus collection of individuals), but also by partitioning the group in various ways to achieve a more nuanced execution of the former (e.g. to carefully minimize exposure to some individuals acting in various capacities at different times, spreading out what liability there *could be* tied to individuals). But this latter practice is not limited to "agent-when groups." Next, we will consider this problem of preferential partitioning.

"Attitudinal Gerrymandering"

One of the things LP allow for in their account of group agency, is for a cognitive division of

labor among subsets of the group agents.[57] For example, individual experts on subject x can deliberate and vote on proposition P while experts on subject y can deliberate and vote on proposition Q.

The duty of making up such caucuses may fall on one or a few individuals. In other words, it will be *someone's* job to divide up the group agent into subgroups. In such a case, a devious individual in this administrative role may, especially with knowledge of how the assignees would vote, carve up the pie in a way that benefits that administrator, by raising the chances of the administrator's preferred group-level output. The proposed analogy is with incumbent politicians who carve a municipal map into districts such that they can with a higher degree of probability ensure their victory (e.g. in a winner-take-all system, redrawing maps such that a high density of supporters currently in one district subsequently span three districts).

But what if the division of labor is *not* the job of one or a few individuals? What if carving up the pie is itself something the group votes on? With such a procedure, attitudinal gerrymandering may well be prevented or minimized. However, it is still a possibility, and therefore still a worry, due to individuals who might vote strategically – a general problem of which this case is one possible instance.[58] Next, we will consider the problem of strategic voting.

Strategic Voting

Members of a group agent that participate in group deliberations are themselves agents with interests. An individual agent might naturally (i.e. as a self-interested, rational individual actor) factor into one's own plans knowledge of the group's decision procedure, in order to procure desirable outcomes for oneself. Considering that group deliberations and "straw poll" voting is often recursive, such an individual gets a peek at the group-level output (e.g. group decision) and can adjust his input (e.g. vote on the motion) in order to affect the outcome that individual desires. In such a case, the individual's participation in the group is second to the individual's ulterior motive – that of controlling the group level output. Thus, the individual contribution (e.g. a vote) is instrumental from the perspective of the individual, and the individual is committed to the group level outcome over performing the individual's role as

[57] List and Pettit, *Group Agency*, 56-57.
[58] A second lesser issue is that group agents may be reluctant to introduce group voting on an issue that was introduced to lighten the group load (i.e. a division of labor). So, it is only to the extent that groups are willing to add group level tasks to an intended task-reducing system that group decisions about subgroup divisions may be the solution.

member, i.e. contributing faithfully, e.g. voting one's genuine view, or honestly and consistently deliberating from the group's point of view when deliberating via group discussion.

An agent can steer the group one way in group discussions by withholding information or exaggerating facts on subjects of which the individual is relied upon by the group to be an expert. With clearer iterated voting procedures, an individual may depart from their role as group member and engage in strategic voting – voting as described above with the intended effect to sway the outcome.

This can be demonstrated using LP's favored type of model. An individual agent with knowledge of the decision procedure (which encompasses background knowledge regarding the logical relations between propositions) and perhaps knowledge of how one's peers will vote, may take up a strategic and exploitative plan to "input" into the group disingenuous votes on propositions with regard only to the desired group output and not the propositions themselves (given the wholistic nature of judgement aggregation on LP's account).

Summary of Problems for LP

Part of the motivation for LP's (and others') account of group agency, is that if we do not see some groups as real agents in society, we miss out on the best and often most relevant descriptions of the world. But we can flip this concern around. In each of the above cases, we could equally miss out on the significance of events in the world if we try to explain them at the group level. In other words, just as we would miss the significance of some events at a higher level if we didn't recognize the agency of groups, so might we miss the right explanations if we incorrectly attribute to groups what should be attributed to individuals.

Strategic voting undermines attempts to exploit the power of collective cognition in decision making. Agent-when groups violate group-level rationality. Attitudinal gerrymandering allows parasitic individuals to pull the strings of a group agent like a puppet. In all three cases the result is a group that risks losing intelligibility as an agent. In such cases of disunity, we could not make sense of group decisions and actions at a group level. A supposed group agent that does not act as we would expect an agent to act is not unlike a mentally disturbed individual – the lack of consistency and coherence undermines the value of treating and considering them as agents, and to continue to do so invites chaos, or at the least, is not at all helpful.

Before proceeding, we might at least raise the possibility of adopting an "exclusive view" of agency: *either* individuals within a group are agents *or* the group is an agent. In other words, we might at least question whether unity could obtain at a level in which an agent has agents as its parts. If we doubt the feasibility of nested agents, we might then have reason to adopt an exclusive view about multi-level agency in which nested individuals or systems can never be nested *agents*.[59] On this approach, the problems for potential disunity discussed above are not problems for group agents or group agent accounts, but reasons for a theoretical commitment to exclusive agency.

So then, we could make a distinction between inclusive and exclusive views of agency in which only the latter adheres to an anti-nesting principle and endorse the latter. However, to do so signs us up for at least some of the bizarre implications of Rovane's account (e.g. committing to the view that individuals, while acting as group agent members, are not at the same time individual agents - see section 2.2.2). This is one area I see as promising for future exploration (see 6.2.5.). There is indeed a lot to work through here. For example, would our adoption of exclusive agency be based on practical concerns (i.e. feasibility of agents attaining inter-agential unity while themselves remaining agents) or in-principle concerns (i.e., to have agents as one's parts is to *not be* an agent)?

So then, one way to diffuse the above problems for group agents and accounts of group agency (e.g. strategic voting, etc.) is to adopt an exclusive view of agency. On this view, the groups are not agents anyway (or if not, then the human members aren't!), so whatever group-level problem exists, they are not problems for group *agents*. But it is also worth asking whether we can rehabilitate the "inclusive" account (in which both can be agents). Suppose we grant an inclusive account like that of LP. How should we think of these phenomena of cheaters and hijackers? In the case of a mentally disturbed individual, we often look inward for a cause, e.g. a neuropathology or psychological disorder. This is consistent with the algorithm in which we seek the appropriate explanations by descending the hierarchy of Dennett's "stances." We troubleshoot (e.g. taking a "design stance" or "physical stance") with the intention of pinpointing the source of this disunity at a lower level. However, we (usually) do not expect to find a homunculus (agent) to be the cause of intra-

[59] On the subject of consciousness (not agency), Putnam asserts a similar position that Kammerer develops (which is itself a response to Schwitzgebel): an "anti-nesting" principle. See Eric Schwitzgebel, "Is the United States phenomenally conscious? Reply to Kammerer," *Philosophia* 44, (2016): 877–883.

personal disunity (this point will be developed in subsequent subsections).

Similarly, when things go wrong with group agents, we likewise look inward to diagnose the problem. When individuals disrupt group level unity, group-level explanations fail, and we must therefore seek explanations at a lower level. However, unlike the cause of a mentally disturbed individual human, when diagnosing the disunity of a group agent, we *do* expect to potentially discover the root of the problem in the agency of a group member (i.e. an individual agent that makes up the group agent). We shift our intentional gaze down one level of the hierarchy to the individual level. In cases like those above, we discover the root of the group's disunity to be the subversive individual pursuing their own agenda from their own practical, self-interested perspective.

5.2.4 Multi-level Intentional Stance

Gerrymandering, strategic voting, and agent-when groups are potential problems for LP.[60] Group agents will face ongoing problems structuring an organization to prevent such instances. However, retrospectively, we are likely to get to the bottom of these problems (if we do) because of the fact that the defectors involved (the cheaters, the non-cooperators) are literal agents - specifically, paradigm, human agents. When something goes wrong at the group level, we, as individuals, naturally look to peer agents at our level. We have evolved to detect cheaters.[61] We, as humans, are all ultimately individualists. We can't help but value and look for explanations at the individual human level (methodological individualism), because it is individuals that matter most to us (normative individualism) and also because each of us sees the world from the perspective of an individual.[62]

Yet when our agential heuristics about biological entities fail, we 1) may not see the entity as a "group," such that it is intuitive to seek a reductive explanation in terms of "individuals";

[60] Firstly, I do not offer a solution to these problems. Rather I aim to show that identifying these problems alone helps inform our biological endeavors. Secondly, if *strictly* required, the agency criteria these situations violate makes group agency implausible – thus potentially not just a problem for groups, but *accounts* of group agency.
[61] Leda Cosmides, and John Tooby, "Neurocognitive adaptations designed for social exchange," *The handbook of evolutionary psychology* (2005): 584-627; Joseph Henrich and Robert Boyd, "Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas," *Journal of theoretical biology* 208, no. 1 (2001): 79-89.
[62] Saying this does not undermine LP's motivation for an account of group agency. Firstly, groups *and* the individuals that comprise them may be held accountable. Secondly, LP are particularly concerned with responsibility gaps – cases in which no group member(s) can be held responsible, and without a group agent, no agent would be held responsible.

and 2) we are more likely to take a troubleshooting perspective (e.g. Dennett's "physical stance" or perhaps "design stance"). With group agents it is natural to remain in the intentional stance but shift levels (from group down to individual). Why? Because we understand non-cooperators as literal cheaters – individual defectors are acting out of self-interest, at odds with the interest of the group. We can take up the cheater's rational point of view as one agent in competition with others, in contrast with the group's point of view. We are not likely to describe the disunity of a group agent in terms of group-level dysfunction and leave it at that – there will be a story to tell involving lower level interagency.

In biology, taking a multi-level intentional stance does not come as naturally to us. In fact, we are taught that this is a mistake – we should not be anthropomorphizing biological entities. We should not project minds into biological creatures, nor think of them as having goals and purposes, nor should we take evolution to be a directional, purposeful process. Yet in this thesis I hold, as others have (see chapter 4), that so long as we are careful to not always take our intentional attributions literally, agential thinking is a useful tool. A familiarity with the concerns raised in the group agency literature allows us to sharpen those tools (hence, the *multi-level* intentional stance in particular).

To come back to the point, instead of assuming *only* that something "went wrong" at the focal level we are epistemically privileging, we can also *be on the lookout for non-cooperating agents* at a different level. That is, instead of adopting a troubleshooting perspective and looking for broken or malfunctioning parts, we can look for defectors[63] – agents acting rationally in their interest, and thereby thwarting the interests of other agents, e.g. agents that they partially comprise. Sure, the higher-level individual is in some sense "broken." But it is not broken like an artifact, it is broken like a group agent. The way we reason about the latter is entirely different from how we reason about the former. It would be a mistake to commit to providing an explanation of biological phenomena solely at the focal level (e.g. the level of the organism), when the real story is about multi-level agency. Framing this story in terms of individuality alone can lead us astray. Traits exhibited "by" an individual at one level will sometimes be the strategies of individuals at another level. A horse *with cancer* may still be *one* physiological individual, but not because it exhibits the

[63] Though not only defectors, cooperators also (i.e. where you would otherwise expect competition). And not just "up and down" but also "side to side" (see section 5.2.5 on reciprocal altruism).

unity of a single agent, and cancer is not the horse's strategy to maximize fitness. Further, to say only that there is now more than one, where formerly there was one (evolutionary) individual, leaves out the concepts that provide the clearest understanding (e.g. goals, interests, cooperation, defection, competition, reciprocation).

Here we are arriving at a key point, and so I will briefly recap. With a biological focus on troubleshooting, we will be looking for "broken parts" at Level 2 instead of subversive agents at Level 1. We are often right to look to entities or phenomena at Level 1 to explain the problem at Level 2 – e.g., we are not mistaken that cancer cells are maladaptive for the organism. But we miss the full explanatory story if we fully epistemically privilege the Level 2 individual – if we think of Level 1 only in terms of parts or pieces of Level 2. The mistake is to underemphasize (or worse, overlook) the part of the explanation in which we understand that what is bad for the organism is good for the cancerous cell.[64] This understanding is about diverging *interests*.

Agency helps us tell a fuller explanatory story, granting epistemic priority neither to the higher nor lower level (e.g. the plant nor mitochondria). Our best explanations are not articulated in terms of the presence of individuals, defined and delineated by synchronic "recipes" (i.e. criteria considered in abstraction from what is *happening*). Saying there are two individuals instead of one - or even two individuals where previously there was one - does not tell the full story. Superior explanations capture the divergence of two (incompatible) agendas, two competing evolutionary interests, two agents with different goals where previously there was one. It may be harder to at first see the problem than accept the solution, because we already have preferred ways of talking and thinking, e.g. a horse *with cancer.*[65]

[64] This is a theoretical point, not an empirical one, i.e. not subject to the objection that, in fact, it could be bad for the cancerous cell in the sense that it is destroying the environment on which it depends for survival, i.e. its host, the multicellular organism (although as a second-order footnote, there are exceptions to this, e.g. devil facial tumor disease or DFTD).

[65] Several have talked about cancer in the context of biological individuality, e.g., Leo Buss, *The evolution of individuality;* Steven Frank, *Dynamics of Cancer: Incidence, Inheritance, and Evolution* (Princeton: Princeton University Press, 2007). But see Marta Bertolaso and Anna Maria Dieli, "Cancer and intercellular cooperation," *Royal Society open science* 4, no. 10 (2017): 170470. For a reply, see Christopher Lean and Anya Plutynski, "The evolution of failure: explaining cancer as an evolutionary process," *Biology & Philosophy* 31, no. 1 (2016): 39-5. On the immune system both restraining and promoting cancers, see Pradeu, *The philosophy of immunology.* The scope of my discussion in the thesis does not hang on taking a stand one way or another on *what cancer is.*

So then, many of our considerations of biological individuality might make more sense when we reframe the picture of the living world as a complex system of heterogeneous actors with interests aligned, or not, to various degrees. In the group agency debate, the problem cases for LP's account are at least seen for what they are – described in terms of non-alignment of interests among different agents. But in biology such problems may not even be fully seen, because they have not been conceptualized and described from a stance that best describes what matters to us in our investigations. Philosophers of biology have debated cases in which purported biological individuals are at odds with themselves in terms of degrees or kinds of "individuality." This is a good start. But our best and fullest explanations can only be offered when seeing the various "individuals" (which may contain, or be contained in, or be in cooperation with, other individuals) as agents with goals and interests (i.e. through an "as-if agency" lens), pursing ends in a complex system of inter-agency (and intra-agency!). Interesting cases of biological individuality (such as when the unity of individuality is in question) cannot always adequately be explained *in terms of individuality alone*.

"Multi-level intentional stance" is not a phrase that exists in the moral and political philosophy literature. What I have named "the multi-level intentional stance" is an invisible concept used by philosophers (or whomever) when considering group agents. Taking up this stance is something we do naturally when considering group agents.

This is noteworthy because in this particular context we do not troubleshoot our attributional failures by shifting stances, but rather by shifting levels to which the stance is applied. Dennett originally contrasted the intentional stance with the design stance and the physical stance. The latter means our explanations appeal to physical properties and laws; the design stance presumes that some system is designed to function for some purpose, and our predictions are guided by that fact. There is an implied hierarchy such that the physical stance can be applied to designed systems and intentional systems, but the intentional stance cannot necessarily be applied back down the hierarchy.

We treat some system as an agent, but then if it acts unagent-like, we treat it as an artefact, quick to point out dysfunction. If treating it as an artifact with a design and purpose fails, we treat it as a non-purposive physical system. To descend from the intentional to the design, then from the design to the physical stance may entail also examining, predicting, and explaining in terms of components or parts (i.e. following a reductionist trend). However, we

need not couple descending levels of a compositional hierarchy with our "descent" of stances. That is, while we may intuitively think of shifting stances as also shifting levels, this is not required, and indeed can lead us astray. When we consider group agents, we routinely shift levels without shifting stances.

Applying this stance to philosophy of biology is useful when we view biological individuals to be like group agents.

Objections

Here is an objection: Is there a possibility of "getting stuck" and reading into the world much more agency than is there?[66]

Here is a reply to the objection. A stance is not an ontology. We, as agents, take up a stance. Taking the intentional stance inappropriately can harm just as much as taking it appropriately can help. Both the intentional and design stance import assumptions about purposiveness. If we are in the habit of interpreting a system's behavior as being "for something," it might be difficult to detach ourselves from this way of thinking. Duncker[67] describes "functional fixedness" as observed in a clinical setting, which Pinker describes as when "people get fixated on an object's function and forget its physical makeup."[68] As a result of conceptual rigidity, people sometimes fail to solve simple problems. The fictional character from the eponymous TV serious *MacGyver*[69] was notoriously adept at fluently switching between the design and physical stances, assessing his goal in terms of the functions to be performed, then sourcing materials from objects suited for a different function, and finally repurposing those materials to perform the task needed to e.g., pick a door lock or improvise an explosive device. The rest of us non-MacGyver types may at times get stuck in the design stance in a situation when only taking the physical stance will be useful. It is reasonable to suppose likewise that we may sometimes get stuck in the intentional stance (as children often do, anthropomorphizing the world around them).

---

[66] E.g. "Darwinian Paranoia." See R.C. Francis, *Why men don't ask for directions: the seductions of sociobiology* (Princeton: Princeton University Press, 2004).
[67] Karl Duncker and Lynne S. Lees, "On problem-solving," *Psychological monographs* 58, no. 5 (1945).
[68] Steven Pinker, *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century* (London: Penguin Books, 2015), 71.
[69] *MacGyver*, https://www.imdb.com/title/tt0088559/?ref_=fn_al_tt_2.

My point here is that the intentional stance is not something taken up as a tool regardless of context. It must be applied by an observant, intentional agent (e.g. a philosopher) sensitive to change and able to make inductive inferences as to whether its application (and ongoing application) is suitable. This is especially important in cases where the phenomenon of multi-level agency may call for a *shift of its application.* By shift, I mean cases in which what is required is not that we take up a different *stance* (e.g. intentional or physical), but rather, a different vantage; we still take up the intentional stance, but at a *different or additional level* e.g. not (only) to the system as a whole, but to one of its sub-systems.

Here is a second objection, the reply to which I have been building towards since chapter 3: Adequate explanations (e.g. of disunified individuals) can be given by citing multiple kinds of individuals (pluralism); perhaps agential explanations are sometimes useful, but they are not needed to give the best explanations.

Here is a reply. Neither individuality, nor pluralism about individuality, is sufficient to satisfactorily describe the troubling or interesting cases of biological individuality which concern us most (e.g. individuals to a degree, intra-organismal conflict, interactions between different kinds of individuals). On Pradeu's immunological view (section 3.4.2), cancerous and non-cancerous cells can together comprise a physiological individual - as does a long-term parasite and its host. The term "physiological individual" applies equally to different types of symbiotic relationships so long as there is biochemical continuity (i.e. so long as some entity is tolerated by the immune system). This preserves our intuitions and conceptual preference to speak about a single individual that *has* a parasite or *has* cancer (interchangeable examples because endogenous versus exogenous origin is not relevant for Pradeu's account). We talk of *one having cancer* as opposed to *one becoming two*. But for Pradeu's immunological view, this is contingent on the cancerous cells being tolerated, not rejected (as it is the immune response which delineates individuals).[70]

The analogous case in a group agent would make tolerated corruption and degeneration perpetrated by group members definitional of a group agent (or kind of group agent). This simply dissolves the unity conditions for agency and reconceives group agents as groups. The relevant information is not that the group can still be a group in some sense, but rather that

[70] For a more nuanced discussion, see Thomas Pradeu, *The philosophy of immunology*, 30.

there are (at least) two agents. This is quite natural for our intuitions about people and groups (especially assuming an "inclusive" concept of agency, which I presume most hold, like LP), because the parts of the group are already presumed agents - their own goals and preferences are always presumed and suspect to be potentially out of alignment with group goals and preferences.

We can create categories of "individuals" in various senses to account for intuitions across a range of circumstances just as we can maim our legs to squeeze into different sized britches. In doing so we can at least avoid the problems of ambiguity so often cited as the motivation for the literature on biological individuality. But on an account like Pradeu's, "physiological individual" has the potential to accommodate heterogeneity at the cost of not discriminating between defection and non-defection. We know not from the term "physiological individual" whether incentive compatibility between whole and parts obtains. Some philosophers of biology could respond that this can be captured with further descriptions in terms of individuality, i.e. two evolutionary individuals within, or comprising, one physiological individual. I argue that conceiving of multi-level biological processes in terms of the generation and transmutation of different kinds of individuals is wrongheaded (even if not as ontologically fraught as it sounds, e.g., because we need not be talking about natural kinds).

Biological individuals that exhibit a unity of purpose are like group agents. Biological individuals that tolerate insufficient conditions of unity are like groups approximating agency to a degree. Our agential attributions may work in some cases of the latter, but at other times will fail. Following Okasha, our agential attributions are best guided by conditions as strict as those for group agency. This was discussed in chapter 4. The relevant point for me here is not that we must refrain from thinking of the higher-level entity (e.g. organism) as an agent when disunity is present – although this is important for reasons previously addressed. The point is to acknowledge that *there are* lower level agents! In the context of group agency, this is obvious. In a biological context it is easily overlooked. Asking how many individuals there are tells us only so much; we stipulate what "individual" means, in some sense, before asking this. Asking about agents and their interactions takes us further towards a satisfying explanation of where things have gone wrong. "Physiological individual" leaves us wondering whether this individual exists despite or because of incentive compatibility among lower level agents. "Evolutionary individual" similarly leaves us wondering whether this

individuality is an achievement or degeneration (e.g. a multicellular organism or a defecting cell lineage – one is maintained through unity, the other born of disunity *at that level*).

The full payoff of the ideas raised in this section cannot be entirely made clear until section 5.2.6. For now, let us bring a few threads together. LP's account of group agency requires rational unity to ensure non-redundant, group-level agency. The normative accounts of agency surveyed in chapter 2 emphasize agency as something achieved and maintained. Similarly, multi-level selection theory and the major transitions in evolution show individuality to be an evolutionary achievement (in the sense that individuality is an evolved character). There are a range of biological entities we consider individuals and we do so for different purposes such that some will accordingly argue for a taxonomy of different kinds of individuals. Okasha's Unity of Purpose condition makes visible and suspect the unity of the biological entities we talk and conceive of heuristically in teleological terms. Not all biological assemblages meet this condition. For my purposes, the relevant point is about *why*: many will not because of the traits ("actions," "strategies") of other (purported) biological individuals (e.g. agents within the focal individual). If, in philosophy of biology, we would take up a stance that is germane to our conceptualization of people and groups in the social realm, we see clearly an opportunity to give explanations that feature biological "agents" independent of the individuals they create, comprise, dissolve, inhabit, overlap or interact with. What the group agency literature has to offer philosophers of biology is a multi-level intentional stance.

5.2.5 Before and After, Up and Down, Side to Side, and All Around

The lesson about levels for biology is this: look up and down, but with a stance toward identifying agents. With group agency we are in the habit of looking for defection of intentional agents as a source of group level degeneration or dysfunction. Why? Individual people are crafty, devious, self-interested – being suspicious of such individuals comes quite naturally. However, in the natural sciences we are much more likely to automatically take up a troubleshooting stance when things go wrong (e.g. against our expectations, given our stance) – applying the intentional stance only to the "broken" individual and its maladaptive dysfunctions, looking to its parts in a mechanical fashion. This perspective privileges the focal level in a way that risks missing lower level agency at play. Our investigations might be

better if we are also prepared to shift levels of organization without shifting stances. In short, look up and down!

Following LP's account, loose groups may come about with joint intentions to act on some shared goal. But it is long-term, structured groups that can be *agents*. In the social realm, we are quite used to identifying and tracking the same agent over time. The emphasis is less on the synchronic conditions that constitute an agent, and more on diachronically tracking purported agents over time. Thus, while diffuse collectives are ephemeral, long-lived corporate collectives are genuine agents. Conceptualizing agency in a static way, with definitional criteria, blueprints, snapshots, the presence of certain functional structures, etc. – this will not suffice. In short, we need to look all around – look forward and backward in time and for interactions with other agents, especially given that our individuality concepts are already described relationally in terms of interagency (e.g. cooperation, collaboration, competition, demarcation).

To get the hang of the multi-level, diachronic perspective, let us consider a timeless evolutionary problem.

The Puzzle of Altruism – Explanations at Three Levels

There are different explanations of biological altruism on offer. I will survey these later in this section as a way of demonstrating the need for a multi-level diachronic perspective.

Altruism is a behavior that is costly to oneself but beneficial to another. In biology, the cost and benefits are understood in terms of survival and fitness. Altruistic behavior is prevalent in nature - mothers care for and nurse their young; sentry meerkats make themselves known to dangerous predators when calling to warn conspecifics; worker bees forgo reproduction and raise the queen's young.

Altruism seems to be a genuine puzzle. Natural selection is the differential survival and reproduction of biological entities. We expect that self-interested individuals would be those most likely to pass through the filter of natural selection. Why would one organism act to

benefit another at cost to itself? There are three kinds of explanations offered at three levels in the biological hierarchy: the gene, the organism, and the group.[71]

Because genes have reproductive interests outside of their "host," then it is plausible that altruistic behaviors might be favored by selection, in this case "kin selection." Organisms may, through altruistic behavior, be acting to maximize "inclusive" fitness.[72] W.D. Hamilton demonstrated this mathematically in an elegant and intuitive way with a coefficient of relatedness. For example, if a father shares 1/4 of its genes with its nephews and 1/2 with its own child, a straightforward optimization function renders non-mysterious the case of selection favoring behavior to save 3 of one's nephews over one's own child.[73] The associated notions of kin selection and inclusive fitness are indexed to the level of the gene. Although the terms refer to an entity larger than an individual (i.e. the family), Hamilton is explaining the effects of genes on the behavior of individuals (for the sake of the genes). Altruism at the organism level is self-interest at the gene level.[74]

The notion of "reciprocal altruism,"[75] however, is indexed to the level of the organism. Because reciprocal altruism is about the mutual benefit of two organisms, altruists need not be kin, or even conspecifics. Altruistic behavior may be selected in a case where assisting another results in receiving assistance oneself at a later time. The paradigm case is reciprocal food sharing.[76] There are times when one will fail to find food. There are other times when one will have plenty. Frequently sharing the spoils is a trait that may spread in a population of

---

[71] We need not think of these necessarily as *competing* explanations. There are very many interesting aspects of this yet to be explored, having to do with realism and levels of description (e.g. in metaphysics and philosophy of science). Most notably, many have shown that the theoretical distinctions between kin selection and group selection disappear when modeled mathematically. See e.g., Steven Frank, *Foundations of Social Evolution* (Princeton: Princeton University Press, 1998); Alan Grafen, "Natural selection, kin selection and group selection, in *Behavioural Ecology: an Evolutionary Approach,* 2nd edition, eds., *J.R. Krebs and N.B. Davies* (Blackwell Scientific Publications, Oxford, 1984), 62-84; Stuart West et al., "Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection," *Journal of evolutionary biology* 20, no. 2 (2007): 415-432.

[72] William Hamilton, "The genetical evolution of social behavior," *Journal of theoretical biology* 7, no. 1 (1964): 17-52.

[73] J.B.S. Haldane's famous (maybe apocryphal) words go something like this: "I would lay down my life for two brothers *or* eight cousins." For a similar point see Edward O. Wilson, "Kin Selection as the Key to Altruism: Its Rise and Fall," *Social Research* 72, no. 1 (2005): 160. I use this example to invoke "trolly-type" consequentialist intuitions, but usually altruism examples are about oneself and another (or others), rather than two other parties.

[74] We can debate whether this counts as *real* altruism (as we can for reciprocal altruism, but for different reasons).

[75] Robert L. Trivers, "The evolution of reciprocal altruism," *The Quarterly review of biology* 46, no. 1 (1971): 35-57.

[76] Gerald S. Wilkinson, "Reciprocal food sharing in the vampire bat," *Nature* 308, no. 5955 (1984): 181-184.

self-interested individuals. We can understand this biological mechanism easily enough without recourse to biology by swapping "strategies" for "traits" in the context of an iterated Prisoner's Dilemma (and other types of social choice games in which we can evaluate competing strategies within scenarios of repeated interactions). Indeed, various models and simulations have demonstrated how cooperative strategies triumph over non-cooperation.[77]

Lastly, the notion of group selection is indexed to the level of the group (e.g. colony).[78] Groups containing altruistic individuals will outcompete groups containing selfish individuals. The former groups will proliferate, and the trait of altruism will spread throughout the larger population.[79]

There is an ongoing debate about which of these is the correct or best biological explanation of altruism. For our purposes, this debate is irrelevant. Instead, let us linger a moment to take one of these three, kin selection, and show how agential thinking has not failed us in the case of altruism (i.e. a puzzle at the level of the organism), but rather requires a hierarchical shift in our application of the intentional stance.

In the ubiquitous instances in which we observe an organism assisting conspecifics to its own detriment, Hamilton suggested that such organisms could be acting to further their evolutionary interests through an indirect strategy – that of kin selection.[80] Rather than acting so as to maximize its own fitness, the organism acts to maximize its inclusive fitness. What this means is that the organism's evolutionary interests *include* not only its own fitness, but that of its *kin*. Evolutionary success depends on the spread of one's genes. Those genes are shared by one's relatives. In other words, traits that aid in the spread of one's genes will be favored by natural selection regardless of who carries those genes. Behavior of self-sacrifice then, can be explained by Hamilton's suggestion.

But notice the terms "organism," "acting," and "evolutionary interest" used above. To correctly deploy the theoretical tools of agential thinking, we should not apply the latter two

[77] Robert Axelrod, *The Evolution of Cooperation* (New York: Basic, 2006).
[78] Sober and Wilson, *Unto Others."*
[79] Interestingly, the explanation of how such a group may persist in light of non-cooperating members may be explained in terms of reciprocal altruism.
[80] Just so as not to confuse, the *ta-da* ending to this story is that it is not the *organism's* strategy, as clarified two paragraphs below.

terms to the first. That is, the organism's *behavior* is the *action* of the gene(s) which, via its "lumbering robot,"[81] carries out a strategy in service of the gene's goal: to maximize clones of itself. (The distinction articulated here between "behavior" and "action" is crucial and most salient in cases of parasitism, see below).[82]

Reciprocal altruism and group selection similarly explain altruism by positing the true locus of agency. Kin selection says, "look down;" group selection says, "look up;" reciprocal altruism says, "look side to side." The first two heuristics apply to space, and more specifically, nested hierarchical structure. The latter applies also to time ("before and after" as well as "side to side"), in which noting the repeated or sustained interaction between agents is required to tell the full story.[83]

There is a lot of room to develop more nuanced concepts of nested agency and interagency, and the special context in which the former is a case of the latter, just as with e.g. interactions between corporations and individual humans on LP's account. Perhaps this is one area in which the two bodies of literature might feedback on one another to form a conceptually synergistic "conversation." (I suggest other possible synergies in section 5.3.)

Parasites

So far, we have considered application of the multi-level intentional stance in regard to cancer, male sterility in flowering plants, and altruism. These cases remind us that individuality can be an achievement of unity or the result of disunity.[84] Whether the relevant actors qualify as individuals at any given time depends on our individuality concepts. But the concept of agency allows us to track the relevant entities irrespective of their (dis)unifying role. Cases involving parasites also make good training grounds for deploying our agential heuristics over time, at multiple levels, across species, and throughout bizarre life cycles.

[81] Dawkins' term.
[82] Incidentally, if space allowed, we could develop a more precise technical notion of "action" for biological purposes in line with the unity of purpose, as-if agential heuristic.
[83] Technically all of our explanations will feature time in *some* relevant way, e.g. in different life cycles of cells and the organisms they comprise.
[84] The latter is not unlike Rovane's view of new agents resulting from fragmentation, i.e. fission.

Imagine you are observing three different creatures: a snail (e.g. Ram-Horned snail), a frog (e.g. Green frog), and a bird (e.g. a Heron). You take note of some key features: The bird has a long sharp beak; the frog has an extra pair of severely deformed legs; the snail appears to be without gonads. It is easy to see why the sharp beak benefits the bird. The bird is able to spear its prey. Adaptationist reasoning makes sense in this context. However, the frog's extra legs slow its movements considerably, leaving it vulnerable to predators. The snail, without any capacity to reproduce, is an evolutionary dead-end. It would not be a very useful heuristic to view the frog's crippled legs and the snail's sterility as strategies for enhancing fitness.

Now consider the flatworm parasite, *Ribeiroia ondatrae*, that depends on each of the creatures above to complete its life cycle.[85] The flat worm parasite colonizes within, and feeds off of, the snail's reproductive tissue, reproducing itself asexually (clonally). The snail releases the parasite larvae into the water where tadpoles swim. The parasite larvae interfere with the tadpole's limb development resulting in a frog that is incapable of escaping from birds of prey. The bird eats the frog. A new generation of parasites are produced (sexually this time) within the bird and the cycle begins again when a snail encounters the bird's feces.[86]

Merely writing off the frog's legs as broken, dysfunctional, or maladaptive risks missing something – and indeed, in this case it does. When one individual "uses" another as a means to an end, it makes clear that indexing properties to a single focal individual is unhelpful. In this case, it is quite natural to describe the case in terms of strategies, goals and more particular intentional vocabulary (e.g. "hijack," "steal," "control" etc.).

The full story is best told by tracking agents according to their purposes, inferring that a trait exhibited at one level (extra frog legs) is the strategy of the agent at a lower level (parasite). By analogy, if a group agent is acting in inconsistent ways that we would not expect an agent to act, then it may be that the *behavior* of the group agent is the *strategy* of one of its manipulative members.

[85] Christine Dell'amore, "Parasite Creating Deformed Frogs in Western U.S." *National Geographic,* August 4, 2011. https://www.nationalgeographic.com/news/2011/8/110802-frogs-deformed-parasites-animals-environment-mutants.
[86] The liver fluke (*Dicrocoelium dendriticum)* has a similar tripartite life cycle featuring ants, snails, and cows.

Here are several more similar examples. When crabs reproduce, they produce and protect an egg sack below their abdomen. Parasitic barnacles (e.g. Rhizocephala, *Sacculina carcini*) can prevent crabs from molting and "trick" the crab into caring for its own sac-like reproductive apparatus (externa).[87] Should we think of the crab's behavior of protecting the parasite's reproductive efforts (rather than its own) as a strategy to maximize its own fitness? Of course not. The multi-level intentional stance easily distinguishes a behavior at one level in one individual as the strategy of another individual at a lower level.

Or consider another aspect of this case. When crabs grow, they normally shed their exoskeleton (i.e. molting). If they lose a limb, they can regenerate that limb during the molting process. But these parasitic barnacles can prevent molting. Could we really think that to forgo the capacity to regrow a missing limb is the crab's goal (or strategy towards a goal) – as being in the crab's interest? Of course not. The spotlight of our agential heuristic should shift to that of the parasitic barnacles. Only then can we see that our observations at one level are the strategies of agents acting to fulfill their goals at another level.[88]

The most obvious sign that an organism's agency has been hijacked is in the case of suicidal behaviors. Nematode-like Gordian worms (e.g. *Paragordius tricuspidatus*) infect the brains of grasshoppers (and crickets) causing them to drown themselves.[89] Is drowning in the interest of the grasshopper? Obviously not. Rather, it is in the parasite's interest that its aerial host return to the water so the worm can initiate a new life cycle. To be clear, the point is not that we should be thinking every trait must be explainable in terms of maximizing fitness (risking the wrath of Gould and Lewontin's ghosts). The point is more simply that writing it off merely as maladaptive or dysfunctional at the level of the grasshopper elides the full story – one that includes not only more than one individual, but also more than one agenda.

Dinocamus coccinellae is a wasp that parasitizes lady beetles (*Coleomegilla maculata*). The larva invades, feeds on, and then cocoons externally on the ladybird's leg.[90] Not unlike the

[87] Henrik Glenner & Jens T Høeg, "A scenario for the evolution of the Rhizocephala," in *Modern Approaches to the Study of Crustacea, eds.,* Elva Escobar-Briones and Fernando Alvarez (Boston: Springer, 2002), 301-310.
[88] We can scrutinize whether "levels" is the right framework here. A case is equally made to look "side to side" as it is to look "up and down." An opportunity for future work is to further categorize different kinds of cases. The broader point here is *not the crab.*
[89] Frédéric Thomas, et al., "Do hairworms (Nematomorpha) manipulate the water seeking behaviour of their terrestrial hosts?" *Journal of Evolutionary Biology* 15, no. 3 (2002): 356-361.
[90] Paul Rincon, "Ladybird made into 'zombie' bodyguard by parasitic was," *BBC News,* June 23, 2011. https://www.bbc.com/news/science-environment-13860891; Deborah Braconnier, "A real-life zombie story in

cases of the flatworm and the crab, the parasite's effect is such that the ladybird's behavior becomes that of a bodyguard.[91] The ladybird's twitching behavior that wards off predators is hypothesized to be coerced via the wasp's venom. Remarkably, once the wasp emerges from the cocoon, the ladybird has a 25% chance of recovering from its "zombification."

In all these examples, we have been practicing our ability to switch between levels of agency. Let us now use a final example to put this new skill set to the test. Suppose you observe a carpenter ant departing from its colony. It ascends, and clings to, the underside of a leaf. It remains there until it dies. Well, we might suspect that while its action does not work to further its own fitness, it might somehow further the interest of genes shared with conspecifics, or the colony as a whole. We investigate further and determine this is not the case. It may instead be, we conclude, the work of a parasite. We make further observations, and indeed, confirm this hypothesis. A close inspection reveals that having clung to the leaf, fungal stalks have protruded from the ant's head, releasing spores to the ground below. The fungus *Orphiocordyceps unilateralus* has "hijacked" our ant friend who behaves the way it does not to further its own interest, but the fungus' interests.

Ah, but now, suppose we look closer. The stalks have emerged from the ant's head, but behold, there are no fruiting spores! How can we make sense of this? The ant's behavior is not furthering the interests of its parasite. The answer: A hyper-parasite. The fungus has a fungus. Similar to Okasha's example of the stamen without pollen in flowering plants, to make sense of this case we must descend yet another level to explain sporeless stalks in terms of the "anti-zombie fungus fungus."[92]

Before moving on, let us give some consideration to the sorts of cases that emphasize the diachronic perspective (in addition to the multi-level perspective) required for optimal

---

the life of bugs," *PhysOrg, June 24, 2011.* https://phys.org/news/2011-06-real-life-zombie-story-life-bugs.html; Stephanie Pappas, "The case of the wasp and the zombie ladybug," *MSNBC*, June 22, 2001. https://www.nbcnews.com/id/wbna43487686; Nolween Dheilly et al., "Who is the puppet master? Replication of a parasitic wasp-associated virus correlates with host behaviour manipulation," *Proceedings of the Royal Society B: Biological Sciences* 282, no. 1803 (2015): 20142773; Nicholas Well, "Wasp virus turns ladybugs into zombie babysitter," *Science,* February 10, 2015. https://www.science.org/content/article/wasp-virus-turns-ladybugs-zombie-babysitters-rev2.

[91] Carl Zimmer, "Meet Natures Nightmare Mindsuckers," *National Geographic,* November 30, 2014. https://carlzimmer.com/mindsuckers-meet-natures-nightmare-128.

[92] Sandra Andersen et al., "Disease dynamics in a specialized parasite of ant societies," *PloS one* 7, no. 5 (2012): e36352.

explanations of these sorts of phenomena (e.g., lady beetle and parasitic wasp detailed above). For the purpose of the example, assume we are dealing with a long-term parasite – before, during, and after it occupies its host. An organism which acquires a parasite (going from t1 to t2), behaves under the controlling influence of that parasite (at t2), and then purges the parasite (t3) may all the while be counted as a single organism or physiological individual, and yet we could not plausibly index the various behaviors across these times to a single *agent*. Many of the organism's behaviors at t2 would need to be explained by the parasite and not the host (i.e. the parasite's "goal" driven behavior). Normal predictions and explanations made at the host level would fail, for example, were the host to engage in unnecessary, non-fitness enhancing suicidal behavior at time t2.[93] Yet, such behavior might enhance the fitness of *the parasite*. This is the natural way of explaining things. By contrast, while we can sharpen our uses and senses of "individuality," individuality alone misses the mark. Sure, there was only one physiological individual at t2, but two at t1 and t3. There were two evolutionary individuals throughout t1, t2, and t3.

The lesson should be clear: our best explanations will not come from settling debates about biological individuality. We need (in fact we already use, if implicitly) agential thinking in our reasoning about these things. Many cases of disunity or borderline individuality will be much more complex, featuring many heterogenous aggregations with selection acting at different levels to a degree. Other cases will be less complex. Whatever the case, we will not explain the *vagueness* of biological individuality by attending to the *ambiguity* of biological individuality.[94] Accounts of individuality, even pluralistic accounts, will not suffice for our best explanations. Borderline cases are those for which we are most in need of "individuals" to explain actors at multiple levels. Yet it is individuality that is the issue. Why not use the explanatory tools that come naturally, using relational agential language that does not depend on what we (have yet to) decide about what counts as an individual?

For the subject of group agency (section 5.1), it turned out that the explanations that capture what matters to us were not given in terms of agency, but individuality (and really, human beings). For the subject of biological individuality, it is the explanations that feature agential

[93] A cricket that drowns itself is explicable only when we understand that "part of" the cricket (i.e. the parasite) has an interest and strategy contrary to the entity of which it is a part.
[94] Recall Clarke and Okasha show these to be the two main issues concerning biological individuality (see section 3.2). Ambiguity can be reframed as criterial pluralism and vagueness accommodated with degrees of individuality.

thinking and associated language (e.g. strategies, goals) that captures what distinguishes, and matters to us about, individuals.

5.2.6 Methodological Agentialism

I have been arguing that our best descriptions and explanations will not come from simply defining and delineating biological individuals, but rather by employing an agential perspective able to span lifetimes, interactions, and hierarchical levels. If I have succeeded in showing this, then we can also argue for a prescriptive version of the heading featured above (in section 5.2.5): *Look Side to Side, Up and Down, Before and After, and All Around!* We can make a normative claim about agential reasoning.

What this means is that, where defining and delineating individuals leaves a deficit in our understanding, positing as-if agency offers a better way to index actions to agents and their goals – a mapping that may span hierarchical levels. This will be most useful in cases of observations that confound our expectations about biological phenomena at a focal level. Therefore, we *should* be looking for opportunities to correctly apply the agential perspective to make best sense of, and provide the fullest explanations for, our observations of biological individuals up and down the hierarchy.

I explored the topic of adaptationism in Chapter 4 to provide background for Okasha's work in which he notes that agential thinking is a form of adaptationist reasoning. Recall that Godfrey-Smith distinguished three kinds of adaptationism: Empirical, Explanatory, and Methodological (see section 4.3.1).

I would like to piggyback onto Godfrey-Smith's categories for two reasons: 1). To give an alternative orientation towards talk of taking the intentional stance which might better appeal to philosophers of biology, and 2). Situate my own offering in this section as a continuation of Okasha's.

Just as Godfrey-Smith distinguished three forms of adaptationism, we could talk about three forms of agentialism. Okasha argues very persuasively for what I would call "explanatory agentialism." Okasha draws our attention to observations of the biological world that motivate our tendency to reason agentially: organisms are the locus of goal-directed

activities; they have flexible behavior; they bear adaptations. In other words, like Godfrey-Smith's explanatory adaptationism, Okasha's work highlights the big questions that call for big answers – there is something very important about the as-if agency of biological individuals. The big answer in the case of explanatory adaptationism is natural selection. I would suggest that the big answer in the case of as-if agency is the Unity of Purpose exhibited by biological individuals.

I take my own view to be offering a complementary emphasis on what I would call "methodological agentialism." Having seen the utility of and accepted justification for agential thinking, we can make an argument for *prescribing* the agential heuristic to researchers, especially in the context of the present topic, i.e., telling the full story about biological phenomena that "individuality" alone cannot (stressing that the unity that makes or breaks individuality is the product of inter-agency). Like methodological individualism, methodological agentialism is a normative position. It offers a conditional imperative: if you want to best address those big questions, then wearing agency goggles in the manner prescribed herein makes for a useful approach to your investigations.[95] I believe Godfrey-Smith, who coined the phrase "agential thinking" in the context of expressing his worry about it, provides a worthy premise towards an argument for methodological agentialism in the following passage:

> When we think about agents and agendas, we think differently and more acutely than we do about abstract logical and causal relations. The strategic perspective on evolution is a way of scientifically engaging this high-powered side of our mind.[96]

There are two things that should be made clear. To a large degree, "methodological agentialism" is a rearticulation and rebranding of the intentional stance. This is not an objection (i.e. a criticism about redundancy) but rather a confirmation of the stated aim, i.e., to satisfy the first *desideratum* listed above. "Methodological agentialism" offers an empirical backdrop in which the "levels" to which we would apply our agential gaze correspond with the levels and units of selection. Philosophers of biology may feel more comfortable,

[95] The third form of agentialism not explicitly addressed here would be "empirical agentialism," which would be an empirical claim about how much of the biological world exhibits the unity that warrants as-if agency ascriptions.
[96] Godfrey-Smith, *Darwinian Populations and Natural Selection,* 10.

empowered, and enabled applying the heuristic considered in this framework (as opposed to "leaving" biology to use an "outside" tool – the intentional stance).

Secondly, I think Okasha can best be read as not *solely* an explanatory agentialist – he is also making the case for *methodological* agentialism. I do not want to inflate my own position as especially unique or novel. The thrust of my argument is to emphasize the big "ought" in counter-distinction to the big taboo we discussed in chapter 4. We have come a long way from the old worries about vitalistic intuitions and misapprehensions of evolution as goal-directed and forward-looking. Methodological agentialism says it is time to proceed without tying a hand behind our back. Agential reasoning, on a modern understanding and in a proper context, is justified and useful.

To sum up so far, we have firstly taken from the group agency literature a perspective about agency at different, multiple hierarchical levels. Secondly, we have borrowed the prefixes Godfrey-Smith used to distinguish different forms of adaptationism and applied them to agential thinking. "Methodological agentialism" is a term we can use to stand in for the approach philosophers of biology can learn from engagement with the group agency literature along the lines explored in this section.

For the remainder of this section, I will attempt to tie the various threads of this section together. We can now round out and show more clearly the ideas I developed in sections 5.2.4 and 5.2.5 by once again considering two of the cases raised previously in sections 5.2.1 and 5.2.5. In a case of cancer, we tend to change stances but not levels. We explain cancer in terms of defects but not defectors. This is fine for oncologists. But fuller explanations, especially explanations of individuality as something evolutionarily achieved and ontogenetically maintained, would include the relevant as-if actors involved. There is a population of cells. Then there is a mutation. Now there is variation. The population is now visible to natural selection. Distinct cell lineages compete within the environment that is the multicellular organism. The population of cells no longer work together towards the goal of maximizing the organism's fitness. This disunity points to distinct goals of the parts. The observation of distinct goals licenses our explanation in terms of distinct agents, regardless of what chunks of this phenomenon we dub individuals of this or that kind, at this or that part of the process. When something like a toy or a car breaks, we say: *what happened? What malfunctioned?* But when a group agent acts in some dissociative "broken" way, we say: *who*

messed up? Who cheated? Who got lazy? We do not look for explanations at the same level, but we do remain in the same stance. We don't have to be reminded that the intentional stance is available to us at the level of the parts. Methodological agentialism asks us to cultivate this attitude towards our explanations of biological individuality.

The reason individuality pluralism remains unsatisfactory is due to the contingent factors that constrain even a range of categories. When we take up our explanatory strategy towards a multicellular organism with cancer, it does not make sense to see a tumor as a strategy working in concert with other strategies (other traits) towards maximizing the multicellular organism's fitness. Okasha's point is that agential attributions at the level shown to exhibit disunity will (or may) not be useful. I think it is up for debate what threshold of disunity is required to make this plausible. I see Okasha's contribution here is calling our attention to *other agents* as source(s) of disunity, particularly those at other levels, more particularly, within nested compositional hierarchies. It seems rather distracting to offer an explanation in terms of different individuals (multicellular organism, cancerous cells, normal cells), of different kinds (evolutionary, physiological), at different times (before or after the mutation, whether or not the cancer is tolerated). The better explanatory story is one about different agents, acting towards different goals from distinct rational points of view. Explanations given in terms of (as-if) agents need not, and should not, depend on what does or does not count as a biological individual. On many views a cell within a multicellular organism might only be a "potential individual" (i.e. only a part, so long as all is well). But it can *always* be thought of as an agent – whether it be one with incentives aligned with other agents working towards a single goal, or one with unaligned incentives working towards different goals. The group agency perspective reminds us that individuality is something more or less achieved and maintained and subject to change. Our best explanations don't come solely in terms of delineating and defining individuals. To do so in some cases paints a picture of individuals transmutating or popping in and out of existence.

Similarly, we can again reflect on the case of parasites. We start out by observing the focal entity (i.e. that which is normally presumed *not* to be parasitized) behaving in a way uncharacteristic of an agent (e.g. to its own detriment). In section 5.2.5 we surveyed a number of cases in which we determined that *behavior* at one level was the *action* of an agent at another level. Applied here, methodological agentialism is not a methodology for discovering new cases of parasitism. Rather it is about the application of the heuristic. It shows that we do

not abandon the intentional stance, but rather apply it elsewhere (instead of or in addition to the focal level). Non-adaptive behavior (or odd traits-qua-strategies) is not evidence that the behavior is not a strategy, but it might be evidence that the corresponding agent is not the focal entity in which the behavior was observed. This way of thinking is independent of whether the host, parasite, or both count as individuals of whatever kind at whatever time. Take notice at how we find it precise to use relational language in this case, distinguishing "host" and "parasite," independent of whether or when one or both of these chunks of matter count as an individual at some time.[97] We can tell the stories of individuality (the creation, loss, and maintenance of individuality) in the language of agency, in which agents and their agendas need not map to any candidate individuals.

In closing, I would like to suggest that one take away from my work is a message similar to that of Karen Kovaka.[98] She argues that the quality of a biologist's work does not suffer, at least not as severely as some have argued,[99] just because the debate about biological individuality has yet to be resolved. In other words, the theoretical and practical problems of biological individuality are not as connected as Clarke and others would have us believe. Seeing the debate as a kind of crisis is due to the over emphasis on the idea that theory informs empirical work; Kovaka reminds us that empirical work often informs theory. "We do not need to settle the individuality debate to do good science. Rather, good science may be our best hope for settling the individuality debate."[100] Furthermore, some conceptions of individuality will be better than others. We need not settle the entire debate to steer clear of unhelpful or problematic conceptions. Concerns about biologists talking past each other or mismeasuring evolutionary change due to a lack of clarity on what counts as an individual are exaggerated. Biologists can reason from their observations and document their counting methods clearly. In other words, biologists are perfectly capable of putting on a minimal philosopher's hat by asking of one another, *what do you mean by "individual?"* Besides, even with well-documented counting methods, we should not expect to solve edge cases of individuality: agreement on criteria is not agreement on individuals. There will always be

[97] Note that we tend not to say this in the cancer case. But maybe we should for some explanations! The multicellular organism is always a potential host and its cells always potential individuals.
[98] Kovaka, "Biological individuality and scientific practice."
[99] Pepper and Herron, "Does biology need an organism concept?"622; Clarke, "The multiple realizability of biological individuals," 414; Ruiz-Mirazo et al., "Organisms and their place in biology," 210.
[100] Kovaka, "Biological individuality and scientific practice," 1096.

marginal cases, regardless of whatever criteria we take to be definitive.[101] "Borderline individuals are an ineliminable feature of the biological world."[102] In other words, both the vagueness and ambiguity problems do not hold up the work of biologists.

My offering of an agential perspective makes a similar point, but within the non-empirical domain. Methodological agentialism is a potentially fruitful way to engage with the literature independent of a likely never-to-be-resolved debate on what things count as biological individuals. Our explanations need not be constrained on the basis of theoretical commitments to constitutive qualities, necessary criteria, or verdicts on edge-cases. Methodological agentialism has the potential to cut straight to the salient aspects of the individuality phenomena in which we take interest. We can have a better, fuller conversation about biological individuality by adopting a perspective and using a relational language that comes naturally to us in other domains – if only we give ourselves permission to do so.

5.2.7 Summary

In this section, we have explored how the topic of group agency might meaningfully inform our considerations of biological individuality. The general lesson we have learned is that the concept of individuality is insufficient for providing the fullest, most relevant explanation about biological individuals. Inspired by the unique context of a group agent, in which we might have agents with cross purposes at multiple hierarchical levels, we can take an approach to describing biological individuality in the language of goals, interests, actions, and agents.[103] A more specific lesson is offered in regard to the shifting of the intentional stance, in which we can avoid missing the whole story by considering the *agendas* of individuals, especially in cases when an odd trait at one level can be reevaluated as the action of an agent at a different level. To be a "methodological agentialist" is to take on, and advocate for, this approach.

5.3 **Group Agency Informing Biological Individuality Informing Group Agency…**

[101] Metaphysician Alyssa Ney makes the distinction between "linguistic vagueness" and "metaphysical vagueness" – the latter pointing to a messy world independent of our tractable concepts. Alyssa Ney, *Metaphysics: an introduction* (Oxon: Routledge, 2014), 114.
[102] Kovaka, "Biological individuality and scientific practice," 1099.
[103] To put it in terms of chapter 4 – we are better attending to the "why" questions that cry out for answers.

In this brief and final section in the chapter, I start with a pre-existing case in which thinking in terms of groups of actors can inform the literature on biological individuality. I then add to it - bringing the implications back around to the topic of group agency.

Recall from chapter 3 that during an evolutionary transition in individuality fitness is "exported" up to the new, higher level individual.[104] Particles comprising the collective can no longer survive and reproduce on their own, but only as part of the whole. The collective entity becomes "Darwinized," the particles "de-Darwinized."[105] Michod describes this as "fitness decoupling," when evolutionary individuality is conferred from parts to whole (e.g. descendants of free living single celled organisms surviving and reproducing only as part of a multicellular organism).

The Pareto Principle states that when all the members of a group prefer one thing to another, so does the group.[106] Or alternatively stated as a prohibition, the decision of a group should not be contrary to the decisions of all of its members.[107]

Okasha borrows this well-known principle from social choice theory to supplement an understanding of Michod's "fitness decoupling" during an evolutionary transition in individuality.[108] For a collective of particles to be a unit of selection, violations of the Pareto Principle must be present such that we could justifiably claim that the collective's welfare is not reducible to that of the particles. Okasha makes the point that if a group is to be a "Darwinian unit, with its own evolutionary interests, that are not reducible to the interest of its constituents," then group fitness cannot be measured in the number of individual offspring, but rather in the number of offspring *groups* - only then, could Pareto violations

[104] Richard Michod, "The group covariance effect and fitness trade-offs during evolutionary transitions in individuality," *Proceedings of the National Academy of Sciences* 103, no. 24 (2006): 9113-9117; Richard Michod et al., "Life-history evolution and the origin of multicellularity," *Journal of theoretical Biology* 239, no. 2 (2006): 257-272; Michod and Nedelcu, "On the reorganization of fitness during evolutionary transitions in individuality."
[105] Godfrey-Smith and others use this language of "de-Darwinization."
[106] E.g. Christian List, "Social Choice Theory," *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), ed., Edward N. Zalta. https://plato.stanford.edu/archives/win2013/entries/social-choice.
[107] E.g. Boniface Mbih et al., "Pareto violations of parliamentary voting systems," *Economic Theory* 34, no. 2 (2008): 331-358.
[108] Samir Okasha, "Individuals, groups, fitness and utility: multi-level selection meets social choice theory," *Biology & Philosophy* 24, no. 5 (2009): 561-584.

occur.[109] For example, the somatic cells within a multicellular organism will not outlive the organism itself – their lineage lives on only as part of the next generation of multicellular organisms (and only via the germ cells).

Similarly, a group is only an agent when it has interests, goals, desires, and beliefs of its own, not reducible to the interests, goals, desires and beliefs of its constituent members. LP demonstrate that something can be good for a group that is not good for any of its members. In 1.5.1 we surveyed models of what LP have called "Discursive Dilemmas," which neatly illustrate that groups may hold beliefs that no members hold.

Okasha has, in effect, already borrowed a lesson from the literature on social choice theory and applied it to biology. I believe we can usefully bring Okasha's interdisciplinary insight full circle back to the subject of group agency. Arguably, this insight shores up LP's account. The general principle which shows that group agents are agents in their own right is the same as the general principle that shows that collectives that have undergone an evolutionary transition in individuality are individuals in their own right. For those skeptical of mere theory and apparent metaphor, an empirical correspondence (if not basis) such as that highlighted by Okasha arguably lends support to LP's account and gives weight to their claim that group agents are "non-redundant" (as few would view, e.g. the concept of an organism, to be a redundant classification of a group of cells). If the decisions, actions, and attitudes of groups could be reduced to those of its members, positing agency at the group level would be redundant. But "unit of selection" is not a redundant classification. This non-redundancy has an empirical basis informed by observation and experiment.

In both domains in which it is applied, the Pareto Principle provides an operational approach to establishing the independence ("decoupling") of a group from its constituent individuals. In other words, contraventions of the Pareto Principle are evidence that the whole is more than the sum of its parts.

To conclude with some commentary, I would suggest that the example in this section demonstrates a kind of *conversation* between the literatures on group agency and biological

[109] Okasha, "Individuals, groups, fitness and utility," 580. Okasha aims to make the point that the concept of group fitness need not be cashed out in terms of average or total individual fitness, which he explores in detail in later work on multi-level selection. See Okasha, *Evolution and the Levels of Selection*.

individuality. We have considered the Pareto Principle in its social choice context, brought it into a biological context, and then exported it back into the original context, relating it to group agency. This conversation is exemplary of a fruitful, mutually informative feedback loop. One of the main thrusts of this thesis is to advocate for the further exploration of common ground between these topics in the hopes of starting many more such conversations.

5.4 Conclusion

In this chapter I have argued that the topics of biological individuality and group agency might be mutually informative. We saw that the kinds of nuanced categories germane to biology provide a clue for how we might better categorize kinds of individuals and groups within society according to our purposes and what we think matters. We have seen how considering the threats to the unity of group agents can provide a framework for considering biological individuality in terms of agents and their interests, strategies, and goals across time and hierarchical levels.

Chapter 6
Looking Ahead

6.1 Summary

This thesis has explored the topics of group agency and biological individuality. In the previous chapters I have surveyed both literatures focusing on some particular philosophers and accounts (e.g. List and Pettit, Okasha, Godfrey-Smith). I have argued that these two literatures might usefully inform and inspire one another. The debate on biological individuality could benefit from a particular agential perspective that is developed more fully in the literature on group agency. The group agency literature could benefit from the more pluralistic approach taken within the biological individuality debate.

In Chapter 1, I introduced List and Pettit's account of group agency, emphasizing the attention given to group level rationality. In Chapter 2, we considered the topic of agency more broadly. We contrasted LP's conception of agency with those of Korsgaard and Rovane, highlighting the notion that agency is an achievement of unity. I also introduced Collins' work categorizing collectives according to their degree of integration. In Chapter 3, we surveyed some of the literature on biological individuality, laying out the common problems and the recent trend towards a pluralistic approach. In Chapter 4, I argued for the acceptability and utility of (judicious) use of agential language, conceptualizing both agency and individuality as achievements of unity in which the parts work together for the benefit of the whole. In Chapter 5, I argued that the biological individuality debate may usefully inform the group agency debate, and vice versa. Some philosophers of biology consider individuals of different kinds and individuality as a matter of degree; I argued that taking this approach to group agency might generate categories that better capture the details we care about according to our purposes. Philosophers who think about group agents naturally take an intentional stance towards the whole and the parts of those agents; I argued that taking this approach to biological individuality allows us to give fuller explanations about the cases that trouble or interest us most.

6.2 Outstanding Problems and Topics for Future Development

Below are five sketches of topics for future work. The reader will note that while I have not argued for these ideas within the thesis, I have held them in mind and referred to them multiple times. The sketches that follow are all (at least somewhat) related in my view and thus might best be developed all together, or in pairs. On the other hand, these are big ambitious ideas, the arguments for which may need to be carefully constructed in pieces. I aim to leave these details open for now.

6.2.1 Incentive Alignment

In this thesis, we noted that potential problems that arise due to misaligned incentives between individuals and the group agents they comprise could serve as an inspiration for participants in the individuality debate in philosophy of biology to maintain the intentional stance in looking for lower-level selfish *actors* rather than merely troubleshooting by looking for malfunctioning *parts*.

However, the challenge of incentive compatibility might be explored in other ways. LP's account is not sufficient to fully explain the many degenerative cases that might result from self-interested individuals with ends that are misaligned with those of the groups they constitute. It is my view that LP's account, as currently formulated, lacks the resources to capture all the nuances required to meet these challenges. Future work could instead explore other accounts of group agency. For instance, one might use Margaret Gilbert's account, which focuses on the commitments of individuals, to propose other structures and mechanisms that might better be self-enforcing.[1] I find particularly interesting the potential for parallels between how we think about binding agreements when doing political philosophy and evolutionary biology, with a focus on the mechanisms at play during evolutionary transitions. Perhaps much of Michod's work on evolutionary transitions in individuality (referenced in chapters 3 and 5) could be meaningfully related to the notion of agreements, contracts, or promises – specifically those maintained by consequences.

I am interested in comparing the processes in which duties and obligations arise (or are generated) to the evolutionary mechanisms that bind separate entities into a collaborative unity. It might be interesting to develop a technical notion of "solidarity" that draws on David

[1] Margaret Gilbert, *Joint commitment.*

Sloan's Wilson's notion of "shared fate" in evolutionary biology. The idea would be to expand on Clarke's notion of "individuating mechanisms" to develop a sense of solidarity that scales up and down the hierarchy and across domains. This opportunity might best be pursued by first developing Michod's ideas regarding the maintenance of individuality (e.g. "policing mechanisms") in the context of cultural evolution and the project of naturalizing morality.

6.2.2 Alterable and Unalterable Aims

Some accounts of group agency focus on a group with ends supplied by the individuals that form the group specifically with those very ends in mind as the group's purpose. By definition, to alter these ends and the structure that works to achieve them would be to kill the group agent. Other accounts suggest the same group agent can choose and alter its own aims.[2] This may crosscut with another distinction that needs to be made explicit – groups in which all individuals take up the group's point of view when voting and contributing, and those in which the group's view is constructed from the majority (or supermajority or whatever) of individual views in the group. The latter seems suited to an organization that wants to capitalize on the collective cognitive outputs of a group of experts, whereas the former is more concerned with ensuring the established pre-existing group perspective is privileged as a default (e.g. considering a charter, mission statement, creed, etc.). It might be interesting to try to relate these various subsets of types of group agents to cooperative (e.g. fraternal transitions) versus collaborative (in O'Malley's sense, e.g. egalitarian transitions) biological agents.

There may be many promising comparisons at this point with the biological world in the way organisms and species "get organized" evolutionarily and the constraints they inherit. Arguably, groups with alterable aims may be able to adapt easier to a changing landscape (e.g. socially or commercially) at the cost of not being identifiable as the same group over time.

[2] In certain ways this is analogous to, and may benefit from engagement with, the AI literature, in which we entertain the idea of an autonomous computer AI able to edit its own source code (after all, a group agent is a kind of artificial agent).

Here is another avenue to explore. The distinction between X and Y may be valuable for *some* purposes, but not others, e.g. how responsibility is attributed and distributed. For example, in a group agent where individuals vote based on their own views (personal or expert), we might think of the group agent as merely "collective." We might then make efforts to trace responsibility back to individuals, finding it less plausible or desirable that any responsibility be attributed to the group agent. By contrast, in the case of a group agent in which individuals contribute or vote directly from the point of view of the group's pre-established point of view, we may find it more acceptable to attribute responsibility to the group agent, but not necessarily to individuals, who were just playing a role.

One could argue that in the corporate case, we would want to avoid a situation similar to that in which we criticize an actor for the mischievous character he plays, rather than for his portrayal, however masterful, of the character. But of course, individuals could then be called to account for taking on the role. One way to develop this conversation is in a context of a hierarchical power structure in which individuals have different mounts of epistemic access to group planning and group outputs, more or less freedom and autonomy, or more or less power or input within the group. This notion of different "roles" might be meaningfully compared to various forms of symbiosis (particularly asymmetrical forms, e.g. parasitism, "black queens") the evolution of multicellularity, or cooperation in eusocial insects. This subject might also lend itself to a discussion of non-intentional group agents, a possibility briefly mentioned by, but otherwise outside the scope of, LP's work.

6.2.3 Bootstrapping Agency and Agency Begetting Agency

In the biological world, we see individuals realized by groups – groups that are themselves groupings of individuals. Similarly, agency is a social concept that, while describing an individual, only makes sense in relation to other agents. The norms of rational agency are cultivated in an agent by other agents. An agent is born into a community of agents that model and teach the norms that the proto agent is required to uphold. And yet to be an agent, one must meet criteria to be an individual agent, capable of acting as a rationally unified, conversable node in the network. In other words, families (tribes, societies) initiate their young into a system of norms, stewarding their development with instruction and increasing expectations of them for which they are accountable to meet. They then become members of

the community – a group made of others like themselves, namely, reliable, conversable, dependable individual agents.

Some of the groundwork for these ideas was explored in chapter 2. One thesis that might be defended here is analogous to the reality that all living things come from living things, and that the individuals that form groups themselves arise from groups. In this context, it would be interesting to further explore Rovane's metaphysics of agency and try to build tighter analogies between intrapersonal and interpersonal unity (or disunity, i.e. "fracturing"), combined with a look into the work on communitarianism within political philosophy (which stresses how individuals are shaped by societal relations and that recognition of this fact should inform our moral judgements, as well as our policies and design of institutions).

To explore this topic by starting with Rovane's work might proceed as follows. Rovane often repeats that unlike our view of a "natural person," an agent is "always a product of effort and will."[3] There is something paradoxical here, as one may ask: *whose* effort and will? What is the thing that tries to be an agent? If the agency is in the trying (i.e. irreducibly normative commitment), what pre-trying subject (or entity?) serves as agent fodder? (recall the notion of "bootstrapping" used in section 2.3). Presumably the answer is reductionistic, something like, a set meaningfully related intentional episodes arising from a human subject: "according to reductionism, there is no further thing, beyond thinkings and doings themselves, that constitute the point of view of an agent."[4] If this is the answer, it does seem we have discovered a powerful dependence relation of agency (group, multiple or otherwise) on *developing* individual human beings (or Martians, etc.). Rovane says that a rational point of view is not a "metaphysical given," because if so, the "boundaries of the first-person point of view from which it deliberates… would be set by the metaphysical condition of its identity."[5] But of course, many who think of agency differently would happily endorse the latter metaphysical thesis along with its corollaries and consequences: deliberating from the point of view of a group is pretense because a group has no actual (metaphysical) point of view.

This dependence relation can, however, be challenged, and reconstrued somewhat paradoxically if we think the societies (or tribes, or families) that shape the norms of

[3] Rovane, "Is group agency a social phenomenon?" 4874.
[4] Rovane, "Is group agency a social phenomenon?" 4877.
[5] Rovane, "Is group agency a social phenomenon?" 4877.

developing humans are themselves group agents. In section 2.3, I put it this way: “On this view, while group agents depend on individual agents for their existence, individual agents depend on group agents for theirs. Agents come from other agents - individuals from groups, and groups from individuals.” Having said this though, a more fundamental asymmetrical dependence remains: on this planet, there would not be rational agents of any kind were it not for the existence of human beings. This latter point invites a bridging between theorizing about agents and individuals in biology and philosophy.

I could go on at length - there is a range of things we might focus on. What I take to be the most compelling idea is what I described in section 2.3 as the hierarchically symmetrical version of the Collective Achievement Thesis: Group agents are formed by individuals. Individuals are formed by group agents.

6.2.4 One Norm

Considerations of the various aforementioned ideas raise the question of whether agents at different levels are realized in similar or the same ways. Our descriptions of unity are parochial – articulated from the vantage of a particular place in the biological hierarchy, at a certain point in history. Considerations about how evolution itself evolves, and what new individuals arise, open up a larger discussion about major transitions and cultural evolution. That is, we might more seriously explore the evolution of humans and human groups, such that our theorizing about agency, is more directly connected to our theorizing about individuality.

Recall Ellen Clarke’s functionally defined “individuating mechanisms,” which while realizable in many ways, work to maintain individuality. What I will provisionally call the “One Norm Thesis” says that rationality and morality are individuating mechanisms at different levels – “rationality” when applied intrapersonally, “morality” when applied interpersonally. Norms said to be “moral” or “rational” refer based on the context, implicitly picking out which agent something is e.g. “good for.” (“Moral” is used in an unusually broad sense here, generally accommodating a range of interpersonal organizational factors, e.g. rules, laws, taboos, promises, etc.). For example, morality and rationality can equally describe relations between agents (e.g. the former in which individual human beings are the focal agents, the latter when groups are the focal agent).

So then, the One Norm Thesis aims to redescribe morality and rationality more generally, as normativity. This ambitious project may be developed in a number of ways. For example, those who are interested in naturalizing ethics, will seize the opportunity to draw on comparisons with evolutionary transitions and the mechanisms that create and maintain individuals at higher or lower levels. This could be seen as a continuation of the mechanist (as opposed to vitalist) tradition. There is room here to build on my notion of the "multilevel intentional stance," reimagining it as a more literal, not merely heuristic, theoretical tool.

6.2.5 Inclusive versus Exclusive Agency

In chapter 2, I proposed a distinction between inclusive and exclusive views on agency. The former allows agents nested within other agents, while the latter prohibits nesting. Perhaps this last topic, of all the subjects mentioned here in chapter 6, holds the most obvious connection between biological individuality and group agency.

One can argue that individuality must only obtain at one level because natural selection acts only at one level. Multi-level selection theorists would raise objections, insisting that "natural selection may operate simultaneously at more than one hierarchical level."[6] One might grant this and still hold that individuality is located at a single place in the hierarchy, perhaps following Queller and Strassman, picking out the "unit of near unanimous design."[7] One might argue for attributing individuality to the level in which there is most autonomy and control, regardless of evolutionary considerations. Alternatively, one might happily allow that individuals nest within others and remain individuals all the same.

This is yet another opportunity to contrast LP's inclusive view with Rovane's exclusive view of agency. We might have alters within humans within companies within conglomerates within states. There is more room to add to what we have explored in section 5.1 (and specifically what was introduced at the end of section 5.1.5) in regards to different kinds of agents and the more detailed taxonomies we could offer. Yet it remains an interesting and worthwhile project to consider when and whether solely agents in the same sense (or of the

[6] Samir Okasha, "The units and levels of selection," in *A Companion to the Philosophy of Biology*, eds., Sahotra Sarkar and Anya Plutynski. (Malden: Blackwell, 2008): 151.
[7] Queller and Strassman, "Beyond Society," 3143.

same kind) can all be disqualified, in principle, when existing within a compositional hierarchy of two or more levels.

Bibliography

Agar, Nicholas. "On the irrationality of mind-uploading: a reply to Neil Levy." *AI & society* 27, no. 4 (2012): 431-436.

Allport, Floyd H. "Teleonomic description in the study of personality." *Character & Personality; A Quarterly for Psychodiagnostic & Allied Studies* (1937), 202-214.

Amodio, Piero, Johanni Brea, Benjamin G. Farrar, Ljerka Ostojić, and Nicola S. Clayton. "Testing two competing hypotheses for Eurasian jays' caching for the future." *Scientific Reports* 11, no. 1 (2021): 1-15.

Andersson, Jan O. "Evolutionary genomics: Is Buchnera a bacterium or an organelle?." *Current Biology* 10, no. 23 (2000): R866-R868.

Andersen, Sandra B., Matthew Ferrari, Harry C. Evans, Simon L. Elliot, Jacobus J. Boomsma, and David P. Hughes. "Disease dynamics in a specialized parasite of ant societies." *PloS one* 7, no. 5 (2012): e36352.

Antonelli, P. L., L. Bevilacqua, and S. F. Rutz. "Theories and models in symbiogenesis." *Nonlinear Analysis: Real World Applications* 4, no. 5 (2003): 743-753.

Axelrod, Robert M. *The Evolution of Cooperation*. New York: Basic Books, 2006.

Ayala, Francisco J. "Teleological explanations versus teleology." *History and philosophy of the life sciences* 20, no. 1 (1998): 41-50.

Bacharach, Michael. *Beyond individual choice: teams and frames in game theory*. Princeton: Princeton University Press, 2018.

Barrow, John, and Frank Tipler. *The Anthropic Cosmological Principle.* New York: Oxford University Press, 1986.

Barton, Nicholas H., Derek E.G. Briggs, Jonathan A. Eisen, David B. Goldstein and Nipam H. Patel. *Evolution.* Cold Spring Harbor: Laboratory press, 2007.

Beckner, Morton. *The biological way of thought*. University of California Press, 1959.

Bergstrom, Carl T., and Lee Alan Dugatkin. *Evolution.* 1st edition. New York: Norton, 2012.

Bertolaso, Marta, and Anna Maria Dieli. "Cancer and intercellular cooperation." *Royal Society open science* 4, no. 10 (2017): 170470.

Björnsson, Gunnar, and Kendy Hess. "Corporate Crocodile Tears?: On the Reactive Attitudes of Corporations." *Philosophy and Phenomenological Research* 94, no. 2 (2017): 273-298.

Blackburn, Simon. *The Oxford dictionary of philosophy*. Oxford University Press, 2005.

Block, Ned. "On a confusion about a function of consciousness." *Behavioral and brain sciences* 18, no. 2 (1995): 227-247.

Bonner, John Tyler. “Evidence for the sorting out of cells in the development of the cellular slime molds.” *Proceedings of the National Academy of Sciences of the United States of America* 45, no. 3 (1959): 379-84.

Braconnier, Deborah. "A real-life zombie story in the life of bugs." *PhysOrg, June 24, 2011.* https://phys.org/news/2011-06-real-life-zombie-story-life-bugs.html.

Bratman, Michael. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press, 1999.

------. "Précis of shared agency: A planning theory of acting together." *Journal of Social Ontology* 1, no. 1 (2015): 1-5.

Briggs, Rachael. “The normative standing of Group Agents.” *Episteme* 9, no. 3 (2012) 283–291.

Brusca, Richard C., and Matthew R. Gilligan. "Tongue replacement in a marine fish (Lutjanus guttatus) by a parasitic isopod (Crustacea: Isopoda)." *Copeia* 1983, no. 3 (1983): 813-816.

Buss, Leo W. "Evolution, development, and the units of selection." *Proceedings of the National Academy of Sciences* 80, no. 5 (1983): 1387-1391.

------. *The Evolution of Individuality*. Princeton: Princeton University Press, 1987.

Burnet, Frank. *Immunological surveillance.* Oxford: Pergamon Press; 1970.

Chauvier, Stéphane. "Individuality and aggregativity." *Philosophy, Theory, and Practice in Biology* 9, no. 11 (2017): 1-14.

Chiu, Lynn, and Gérard Eberl. "Microorganisms as scaffolds of host individuality: an eco-immunity account of the holobiont." *Biology & Philosophy* 31, no. 6 (2016): 819-837.

Clarke, Ellen. "Is Evolution Fundamental When It Comes to Defining Biological Ontology? Yes." In *Current Controversies in Philosophy of Science*, edited by Shamik Dasgupta, Ravit Dotan, and Brad Weslake, 104-118. New York: Routledge, 2020.

------. "Origins of evolutionary transitions." *Journal of biosciences* 39, no. 2 (2014): 303-317.

------. "Plant individuality: a solution to the demographer’s dilemma." *Biology & Philosophy* 27, no. 3 (2012): 321-361.

------. "The multiple realizability of biological individuals." *The Journal of Philosophy* 110, no. 8 (2013): 413-435.

------. "The problem of biological individuality." *Biological theory* 5, no. 4 (2010): 312-325.

Clarke, Ellen, and Samir Okasha. "Species and organisms: What are the problems." In *From groups to individuals: evolution and emerging individuality*, edited by Frédéric Bouchard and Philippe Huneman, 55-75. Cambridge: MIT Press, 2013.

Collins, Stephanie. "Collective responsibility gaps." *Journal of Business Ethics* 154, no. 4 (2019): 943-954.

Copp, David. "Hobbes on artificial persons and collective actions." *The Philosophical Review* 89, no. 4 (1980): 579-606.

Cosmides, Leda, and John Tooby. "Neurocognitive adaptations designed for social exchange." *The handbook of evolutionary psychology* (2005): 584-627.

Cudd, Ann E. "Individual Agency as Collective Achievement." In *Proceedings of the XXIII World Congress of Philosophy 44*, (201): 5-9.

Cummins, Robert. "Functional Analysis." *The Journal of Philosophy* 72, no. 20 (1975): 741-65.

Damuth, John, and I. Lorraine Heisler. "Alternative formulations of multilevel selection." *Biology and Philosophy* 3, no. 4 (1988): 407-430.

Davidson, Donald. *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.

Dawkins, Richard. *The Extended Phenotype: The Long Reach of the Gene*. New York: Oxford, 1982.

------. *The Selfish Gene. 40th Anniversary edition*. Oxford: Landmark Science, 2016.

------. “The Purpose of Purpose - Richard Dawkins.” Online video clip. *YouTube*, 19 June 2009. Web. 1 October 2019.

Dell’amore, Christine. “Parasite Creating Deformed Frogs in Western U.S.” *National Geographic,* August 4, 2011, https://www.nationalgeographic.com/news/2011/8/110802-frogs-deformed-parasites-animals-environment-mutants.

Denison, R. Ford, and E. Toby Kiers. "Life histories of symbiotic rhizobia and mycorrhizal fungi." *Current Biology* 21, no. 18 (2011): R775-R785.

Dennett, Daniel. *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* New York: Simon & Schuster, 1995.

------. "Intentional systems." *The Journal of Philosophy* 68, no. 4 (1971): 87-106.

------. *The Intentional Stance*. Cambridge: MIT Press. 1987.

Dheilly, Nolwenn M., Fanny Maure, Marc Ravallec, Richard Galinier, Josée Doyon, David Duval, Lucas Leger et al. "Who is the puppet master? Replication of a parasitic wasp-

associated virus correlates with host behaviour manipulation." *Proceedings of the Royal Society B: Biological Sciences* 282, no. 1803 (2015): 20142773.

Dietrich, Franz, and Christian List. "Arrow's theorem in judgment aggregation." *Social Choice and Welfare* 29, no. 1 (2007): 19-33.

DiFrisco, James. "Kinds of Biological Individuals: Sortals, Projectibility, and Selection." The *British Journal for the philosophy of science 70, no. 3* (Sep 2019): 845-875.

DiFrisco, James, and Matteo Mossio. "Diachronic identity in complex life cycles: an organizational perspective." In *Biological Identity: Perspectives from Metaphysics and the Philosophy of Biology*, edited by A.S. Meincke and John Dupré, 177-199. New York, Routledge: 2021.

Dretske, Fred I. *Explaining Behavior: Reasons in a World of Causes.* Cambridge: MIT Press, 1988.

Drucker, Peter F. *Concept of the Corporation*. New York: John Day, 1946.

Duncker, Karl, and Lynne S. Lees. "On problem-solving." *Psychological monographs* 58, no. 5 (1945).

Dupré, John and Maureen O'Malley. "Varieties of Living Things: Life at the Intersection of Lineage and Metabolism." *Philosophy, Theory, and Practice in Biology* 1 (2009): 1-25.

Ereshefsky, Marc, and Makmiller Pedroso. "Rethinking evolutionary individuality." *Proceedings of the National Academy of Sciences* 112, no. 33 (2015): 10126-10132.

Feldhaar, Heike. "Bacterial symbionts as mediators of ecologically important traits of insect hosts." *Ecological Entomology* 36, no. 5 (2011): 533-543.

Folse III, Henri J., and Joan Roughgarden. "What is an individual organism? A multilevel selection perspective." *The Quarterly review of biology* 85, no. 4 (2010): 447-472.

Francis, R.C. *Why men don't ask for directions: the seductions of sociobiology*. Princeton: Princeton University Press, 2004.

Frank, Steven. *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. Princeton: Princeton University Press, 2007.

------. *Foundations of Social Evolution*. Princeton: Princeton University Press, 1998.

------. "The evolutionary dynamics of cytoplasmic male sterility." *The American Naturalist* 133, no. 3 (1989): 345-376.

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66, no. 23 (1969): 829–839.

French, Peter. *Collective and corporate responsibility*. New York: Columbia University Press, 1984.

------. *Corporate Ethics. N*ew York: Wadsworth, 1994.

------. "The corporation as a moral person." *American Philosophical Quarterly* 16, no. 3 (1979): 207-215.

Futuyma, Douglas J. *Evolution*. 2nd edition. Sunderland: Sinauer Associates, 2009.

Garthoff. "Decomposing legal personhood." *Journal of Business Ethics* 154, no. 4 (2019): 967-974.

Ghiselin, Michael T. "A radical solution to the species problem." *Systematic Biology* 23, no. 4 (1974): 536-544.

Gilbert, Margaret. *Joint commitment: How we make the social world*. Oxford University Press, 2013.

Gilbert, Scott F., Jan Sapp, and Alfred I. Tauber. "A symbiotic view of life: we have never been individuals." *The Quarterly review of biology* 87, no. 4 (2012): 325-341.

Godfrey-Smith, Peter. *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press, 1998.

------. “Darwinian Individuals.” In *From groups to individuals: evolution and emerging individuality*, edited by Frédéric Bouchard and Philippe Huneman, 17-36. Cambridge: MIT Press, 2013.

------. *Darwinian populations and natural selection*. Oxford: Oxford University Press, 2009.

------. "Individuality, subjectivity, and minimal cognition." *Biology & Philosophy* 31, no. 6 (2016): 775-796.

------. *Philosophy of Biology*. Princeton University Press, 2013.

------. "Three kinds of adaptationism." In *Adaptationism and Optimality*, edited by S. Orzack and E. Sober. 335-357. Cambridge University Press, 2001.

Goodpaster, Kenneth E., and John B. Matthews. "Can a corporation have a conscience." *Harvard business review* 60, no. 1 (1982): 132-141.

Gordon-Roth, Jessica. "Locke on Personal Identity." *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), edited by Edward N. Zalta, https://plato.stanford.edu/archives/spr2020/entries/locke-personal-identity.

Gould, Stephen Jay, and Richard C. Lewontin. "The Spandrels of San Marco and the Panglossian Paradigm: a critique of the adaptationist programme." *Conceptual issues in evolutionary biology* 2 (1994): 73-90.

Grafen, Alan. “Natural selection, kin selection and group selection. In *Behavioural Ecology: an Evolutionary Approach,* 2nd edition, edited by *J.R. Krebs and N.B. Davies*, 62-84. Blackwell Scientific Publications, Oxford, 1984.

------. "Optimization of inclusive fitness." *Journal of theoretical Biology* 238, no. 3 (2006): 541-563.

------. "The formal darwinism project in outline." *Biology & Philosophy* 29, no. 2 (2014): 155-174.

Grene, Marjorie. “Aristotle and Modern Biology.” In *Topics in the Philosophy of Biology*, edited by Marjorie Grene and Everett Mendelson. Dordrecht: Reidel,1976.

Griesemer, James. "Development, culture, and the units of inheritance." *Philosophy of science* 67, no. 3 (2000): S348-S368.

------. "Reproduction in complex life cycles: toward a developmental reaction norms perspective." *Philosophy of Science* 83, no. 5 (2016): 803-815.

Grissa, Ibtissem, Gilles Vergnaud, and Christine Pourcel. "The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats." *BMC bioinformatics* 8, no. 1 (2007): 1-10.

Haber, Matthew. “Colonies are individuals: revisiting the superorganism revival.” In *From groups to individuals: evolution and emerging individuality*, edited by Frédéric Bouchard and Philippe Huneman, 195-217. Cambridge: MIT Press, 2013.

Haig, David. "Proximate and ultimate causes: how come? and what for?" *Biology & Philosophy* 28, no. 5 (2013): 781-786.

Haksar, Vinit. “Moral agents.” In *Routledge Encyclopedia of Philosophy,* Taylor and Francis, 1998. https://www.rep.routledge.com/articles/thematic/moral-agents/v-1/sections/agents-versus-recipients.

Hamilton, William D. "Innate social aptitudes of man: an approach from evolutionary genetics." *Biosocial anthropology* 53 (1975): 133-155.

------. "The genetical evolution of social behaviour.” *Journal of theoretical biology* 7, no. 1 (1964): 17-52.

Hamrick, J. T., Farhang Rouhi, Arghya Mukherjee, Amir Feder, Neil Gandal, Tyler Moore, and Marie Vasek. "An examination of the cryptocurrency pump-and-dump ecosystem." *Information Processing & Management* 58, no. 4 (2021): 102506, 1-16.

Hansen, Susse Kirkelund, Paul B. Rainey, Janus AJ Haagensen, and Søren Molin. "Evolution of species interactions in a biofilm community." *Nature* 445, no. 7127 (2007): 533-536.

Harper, John L. *Population Biology of Plants*. New York: Academic Press, 1977.

Hartshorne, Charles. *Collected Papers of Charles Sanders Peirce*, edited by Paul Weiss and Arthur W. Burks. Harvard University Press, 1958.

Hasnas, John. "Should corporations have the right to vote? A paradox in the theory of corporate moral agency." *Journal of business ethics* 150, no. 3 (2018): 657-670.

------. "The Phantom Menace of the Responsibility Deficit." In *The Moral Responsibility of Firm, e*dited by. Eric Orts and N. Craig Smith, 89-105. Oxford: Oxford University Press, 2017.

Henrich, Joseph, and Robert Boyd. "Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas." *Journal of theoretical biology* 208, no. 1 (2001): 79-89.

Herron, Matthew D., Armin Rashidi, Deborah E. Shelton, and William W. Driscoll. "Cellular differentiation and individuality in the 'minor' multicellular taxa." *Biological Reviews* 88, no. 4 (2013): 844-861.

Hess, Kendy M. "If You Tickle Us...": How Corporations can be Moral Agents without being Persons." *Journal of Value Inquiry* 47, no. 3 (2013): 319-335.

Hindriks, Frank. "How autonomous are collective agents? Corporate rights and normative individualism." *Erkenntnis* 79, no. 9 (2014): 1565-1585.

Hobbes, Thomas. *Leviathan: With Selected Variants from the Latin Edition of 1668*, edited by E.M. Curley. Indianapolis: Hackett Publishing, 1994.

Høeg, Henrik and Jens T., and Glenner. Høeg. "A scenario for the evolution of the Rhizocephala." In *Modern Approaches to the Study of Crustacea, edited by* Elva Escobar-Briones and Fernando Alvarez, 301-310. Boston: Springer, 2002.

Hull, David L. "A matter of individuality." *Philosophy of science* 45, no. 3 (1978): 335-360.

------. "Are species really individuals?" *Systematic zoology* 25, no. 2 (1976): 174-191.

------. "Individuality and selection." *Annual review of ecology and systematics* 11, no. 1 (1980): 311-332.

Huxley, Julian. *The Individual in the Animal Kingdom*. New York: University Press, 1912.

Janzen, Daniel H. "What are dandelions and aphids?" *The American Naturalist* 111, no. 979 (1977): 586-589.

Jensen, Mark N. "Is Our Group An Agent? Do We Want It To Be?" *Journal of Moral Philosophy* 12, 4 (2015): 539-551.

Jeske, Diane. "Special Obligations." *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), edited by Edward N. Zalta, https://plato.stanford.edu/archives/fall2019/entries/special-obligations.

Johnson, Brian R., and Timothy A. Linksvayer. "Deconstructing the superorganism: social physiology, groundplans, and sociogenomics." *The Quarterly Review of Biology* 85, no. 1 (2010): 57-79.

Jones, Robert C. “Science, sentience, and animal welfare.” *Biology & Philosophy* 28, no. 1 (2013): 1-30.

Kacelnik, Alex. “Meanings of rationality.” In Rational Animals? Edited by S.E. Hurley and M.E. Nudds, 87–106. Oxford: Oxford University Press, 2006.

Kammerer, François. “How a materialist can deny that the United States is probably conscious – Response to Schwitzgebel.” *Philosophia* 43, (2015): 1047–1057.

Kant, Immanuel. *Grundlegung zur Metaphysik der Sitten*, translated with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; reprint New York: Harper & Row, 1964 (originally 1785).

Kennett, Jeanette, and Steve Matthews. "The unity and disunity of agency." *Philosophy, Psychiatry, & Psychology* 10, no. 4 (2003): 305-312.

Killoren, David, and Bekka Williams. "Group agency and overdetermination." *Ethical Theory and Moral Practice* 16, no. 2 (2013): 295-307.

Korsgaard, Christine M. "Personal identity and the unity of agency: A Kantian response to Parfit." *Philosophy & Public Affairs* 18, no. 2(1989): 101-132.

------. *Self-constitution: Agency, identity, and integrity*. Oxford: Oxford University Press, 2009.

------. "The Normative Constitution of Agency." In *Rational and Social* 190-214. Oxford: Oxford University Press, 2014.

Kornhauser, Lewis A., and Lawrence G. Sager. "The One and the Many: Adjudication in Collegial Courts." *California Law Review* 81, no. 1 (1993): 1-59.

Kovaka, Karen. "Biological individuality and scientific practice." *Philosophy of Science* 82, no. 5 (2015): 1092-1103.

Lean, Christopher, and Anya Plutynski A. “The evolution of failure: explaining cancer as an evolutionary process.” *Biology & Philosophy* 31 no. 1 (2016): 39–57.

Lewontin, Richard C. "The units of selection." *Annual review of ecology and systematics* 1, no. 1 (1970): 1-18.

Lidgard, Scott, and Lynn K. Nyhart, eds. *Biological individuality: Integrating scientific, philosophical, and historical perspectives*. University of Chicago Press, 2017.

List, Christian. "Social Choice Theory." *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2013/entries/social-choice.

------. "What is it Like to be a Group Agent?" *Noûs* 52, no. 2 (2018): 295-319.

List, Christian, and Peter Menzies. "Nonreductive physicalism and the limits of the exclusion principle." *The Journal of Philosophy* 106, no. 9 (2009): 475-502.

List, Christian, and Philip Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents.* Oxford: Oxford University Press, 2011.

Lovejoy, Arthur. *The Great Chain of Being a Study of the History of an Idea: The William James Lectures Delivered at Harvard University, 1933*. Cambridge: Harvard University Press, 1964.

Lloyd, Elizabeth. "Holobionts as units of selection: holobionts as interactors, reproducers, and manifestors of adaptation. In *Landscapes of Collectivity in the Life Sciences,* edited by S. B. Gissis, E. Lamm and A. Shavit, 351– 368. London: MIT Press, London, 2017.

Lloyd, Elisabeth, and Michael J. Wade. "Criteria for holobionts from community genetics." *Biological Theory* 14, no. 3 (2019): 151-170.

Manning, Rita C. "Corporate responsibility and corporate personhood." *Journal of Business Ethics* 3, no. 1 (1984): 77-84

Margulis, Lynn. "Words as battle cries: symbiogenesis and the new field of endocytobiology." *Bioscience* 40, no. 9 (1990): 673-677.

Margulis, Lynn and Dorion Sagan. "What is Life?" New York: Simon and Schuster: 1995.

Mariscal, Carlos, and W. Ford Doolittle. "Life and life only: a radical alternative to life definitionism." *Synthese* 197, no. 7 (2020): 2975-2989.

May, Larry. *Sharing responsibility*. Chicago: University of Chicago Press, 1992.

Maynard Smith, John. *Evolution and the Theory of Games*. New York: Cambridge university press, 1982.

John Maynard Smith, John and Eors Szathmary. *The Major Transitions in Evolution*. New York: Freeman 1995.

Mayr, Ernst. "Cause and effect in biology." *Science* 134, no. 3489 (1961): 1501-1506.

------. "Teleological and Teleonomic, a New Analysis." *Boston Studies in the Philosophy of Science* 14 (1974): 133–159.

------. "The idea of teleology," *Journal of the History of Ideas* 53, no. 1 (1992): 117-135.

------. "The multiple meanings of teleological." In Towards a new philosophy of biology: Observations of an evolutionist, edited by Ernst Mayr, 38-66. Harvard University Press, 1988.

------. *What Evolution Is*. New York: Basic Books, 2001.

Mbih, Boniface, Issofa Moyouwou, and Jérémy Picot. "Pareto violations of parliamentary voting systems." *Economic Theory* 34, no. 2 (2008): 331-358.

McShea, Daniel W. "The hierarchical structure of organisms: a scale and documentation of a trend in the maximum." *Paleobiology* 27, no. 2 (2001): 405-423.

Mercier, Hugo, and Dan Sperber. *The enigma of reason*. Harvard University Press, 2017.

Michod, Richard. *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality.* Princeton: Princeton University Press, 1999.

------. "Evolution of individuality during the transition from unicellular to multicellular life." *Proceedings of the National Academy of Sciences* 104, no. S1 (2007): 8613-8618.

------. "The group covariance effect and fitness trade-offs during evolutionary transitions in individuality." *Proceedings of the National Academy of Sciences* 103, no. 24 (2006): 9113-9117.

Michod, Richard E., and Matthew D. Herron. "Cooperation and conflict during evolutionary transitions in individuality." *Journal of evolutionary biology* 19, no. 5 (2006): 1406-1409.

Michod, Richard, and Aurora M. Nedelcu. "On the reorganization of fitness during evolutionary transitions in individuality." *Integrative and Comparative Biology* 43, no. 1 (2003): 64-73.

Michod, Richard, Yannick Viossat, Cristian A. Solari, Mathilde Hurand, and Aurora M. Nedelcu. "Life-history evolution and the origin of multicellularity." *Journal of theoretical Biology* 239, no. 2 (2006): 257-272.

Miller, Seumas. *Social action: A teleological account*. Cambridge University Press, 2001.

Millikan, Ruth Garrett. "In defense of proper functions." *Philosophy of science* 56, no. 2 (1989): 288-302.

Misselhorn, Catrin. "Collective agency and cooperation in natural and artificial systems." In *Collective agency and cooperation in natural and artificial systems*, edited by Catrin Misselhorn, 3-24. London: Springer, 2015.

Mitton, Jeffry B., and Michael C. Grant. "Genetic variation and the natural history of quaking aspen." *Bioscience* 46, no. 1 (1996): 25-31.

Mölder, Bruno. *Mind Ascribed: an Elaboration and Defence of Interpretivism*. Philadelphia: John Benjamins, 2010.

Molter, Daniel J. "On mycorrhizal individuality." *Biology & Philosophy* 34, no. 5 (2019): 1-16.

Monod, Jacques. *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. London: Collins, 1972.

Moran, Nancy A., and Daniel B. Sloan. "The hologenome concept: helpful or hollow?." *PLoS biology* 13, no. 12 (2015): e1002311.

Moreno, Álvaro, and Matteo Mossio. *Biological Autonomy: A Philosophical and Theoretical Enquiry*. Dordrecht: Springer, 2015.

Neander, Karen. "Biological Function." In *Routledge Encyclopedia of Philosophy,* Taylor and Francis, 2012. https://www.rep.routledge.com/articles/thematic/biological-function/v-1.

------. "Functions as selected effects: The conceptual analyst's defense." *Philosophy of science* 58, no. 2 (1991): 168-184.

Ney, Alyssa. *Metaphysics: an introduction*. Oxon: Routledge, 2014.

O'Malley, Maureen. "Is Evolution Fundamental When It Comes to Defining Biological Ontology? No." In *Current Controversies in Philosophy of Science*, edited by Shamik Dasgupta, Ravit Dotan, and Brad Weslake, 93-104. New York: Routledge, 2020.

Okasha, Samir. *Agents and goals in evolution*. Oxford University Press, 2018.

------. "Biological Ontology and Hierarchical Organization: A Defense of Rank Freedom." In *The Major Transitions in Evolution Revisited*, edited by Brett Calcott and Kim Sterelny, 53-64. Cambridge: MIT Press, 2011.

------. *Evolution and the Levels of Selection* (Oxford University Press, 2006).

------. "Individuals, groups, fitness and utility: multi-level selection meets social choice theory." *Biology & Philosophy* 24, no. 5 (2009): 561-584.

------. "The units and levels of selection." In *A Companion to the Philosophy of Biology*, edited by Sahotra Sarkar and Anya Plutynski, 138-156. Malden: Blackwell, 2008.

Pappas, Stephanie. "The case of the wasp and the zombie ladybug." *MSNBC*, June 22, 2001. https://www.nbcnews.com/id/wbna43487686.

Paul, Laurie Ann. *Transformative experience*. Oxford: Oxford University Press, 2014.

Pepper, John W. and Matthew D. Herron. "Does biology need an organism concept?" *Biological Reviews* 83, no. 4 (2008): 621-627.

Pettit, Philip. "Corporate Agency: The Lesson of the Discursive Dilemma." In *The Routledge Handbook of Collective Intentionality*, edited by Marija Jankovic and Kirk Ludwig, 249-60. New York: Routledge, 2018.

------. "Consciousness Incorporated," *Journal of Social Philosophy* 49 (2018): 12-37.

------. "Group Agents Are Not Expressive, Pragmatic or Theoretical Fictions." *Erkenntnis* 79. no. S9 (2014): 1641-662.

------. "Groups with minds of their own." In *Social epistemology,* edited by Alvin Goldman and Dennis Whitcomb, 242-268. New York: Oxford University Press, 2011.

Pfeiffer, Raymond S. "The central distinction in the theory of corporate moral personhood." *Journal of Business Ethics* 9, no. 6 (1990): 473-480

Pinker, Steven. *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century*. London: Penguin Books, 2015.

Pittendrigh, Colin. "Adaptation, natural selection, and behavior." In *Behavior and evolution,* edited by A. Roe and G. G. Simpson. New Haven: Yale University Press, 1958.

Pradeu, Thomas. "Immunity and the emergence of individuality." In *From groups to individuals: evolution and emerging individuality*, edited by Frédéric Bouchard and Philippe Huneman, 77-96. Cambridge: MIT Press, 2013.

------. "Organisms or biological individuals? Combining physiological and evolutionary individuality." *Biology & Philosophy* 31, no. 6 (2016): 797-817.

------. *The limits of the self: immunology and biological identity*. Oxford: Oxford University Press, 2011.

------. *The Philosophy of Immunology.* Cambridge University Press, 2020.

------. "What Is An Organism? An Immunological Answer." *History and Philosophy of the Life Sciences* 32, no. 2/3 (2010): 247-67

Putnam, Hilary. "Psychological predicates." In *Art, mind, and religion*, edited by W.H. Capitan and D.D. Merrill. Liverpool: University of Pittsburgh Press, 1965.

Queller, David C. "Relatedness and the fraternal major transitions." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 355, no. 1403 (2000): 1647-1655.

Queller, David C., and Joan E. Strassmann. "Beyond society: the evolution of organismality." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, no. 1533 (2009): 3143-3155.

Ratcliff, William C., Johnathon D. Fankhauser, David W. Rogers, Duncan Greig, and Michael Travisano. "Origins of multicellular evolvability in snowflake yeast." *Nature communications* 6, no. 1 (2015): 1-9.

Ratnieks, Francis LW, and P. Kirk Visscher. "Worker policing in the honeybee." *Nature* 342, no. 6251 (1989): 796-797.

Reese, Hayne W. "Teleology and teleonomy in behavior analysis." *The Behavior Analyst* 17, no. 1 (1994): 75-91.

Rincon, Paul. "Ladybird made into 'zombie' bodyguard by parasitic was." *BBC News,* June 23, 2011. https://www.bbc.com/news/science-environment-13860891.

Rosenblueth, Arturo, Norbert Wiener, and Julian Bigelow. "Behavior, purpose and teleology." *Philosophy of science* 10, no. 1 (1943): 18-24.

Rovane, Carol. "Group agency and individualism." *Erkenntnis* 79, no. 9 (2014): 1663-1684

------. "Is group agency a social phenomenon?" *Synthese* 196, no. 12 (2019): 4869-4898.

------. *The bounds of agency*. Princeton: Princeton University Press, 1998.

------. "What is an Agent?." *Synthese* 140, no. 1/2 (2004): 181-198.

Ruiz-Mirazo, Kepa, Arantza Etxeberria, Alvaro Moreno, and Jesús Ibáñez. "Organisms and their place in biology." *Theory in biosciences* 119, no. 3 (2000): 209-233.

Santelices, Bernabé. “How many kinds of individual are there?” *Trends in ecology & evolution* 14, no. 4 (1999): 152-155.

Sapp, Jan. "Paul Buchner (1886–1978) and hereditary symbiosis in insects." *International Microbiology* 5, no. 3 (2002): 145-150.

Schnable, Patrick S., and Roger P. Wise. "The molecular basis of cytoplasmic male sterility and fertility restoration." *Trends in plant science* 3, no. 5 (1998): 175-180.

Schwitzgebel, Eric. If materialism is true, the United States is probably conscious. *Philosophical Studies* 172 (2015): 1697–1721.

------. “Is the United States phenomenally conscious? Reply to Kammerer.” *Philosophia* 44, (2016): 877–883.

Seeley, Thomas D. *The wisdom of the hive: the social physiology of honey bee colonies*. Harvard University Press, 2009.

Short, T. L. "Peirce's Concept of Final Causation." *Transactions of the Charles S. Peirce Society* 17, no. 4 (1981): 369-82.

Singer, Peter. "Discussing Infanticide." *Journal of Medical Ethics* 39, no. 5 (2013): 260.

Sober, Elliott. "Three differences between deliberation." In *Modeling Rationality, Morality, and Evolution,* edited by Peter Danielson, 408-422. Cary: Oxford University Press, 1998.

Sober, Elliott, and David Sloan Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press, 1998.

Smith, Subrena E. "Organisms as persisters." *Philosophy, Theory, and Practice in Biology* 9, no. 14 (2017): 1-16.

Skillings, Derek. "Holobionts and the ecology of organisms: multi-species communities or integrated individuals?" *Biology & Philosophy* 31, no. 6 (2016): 875-892.

Suárez, Javier, and Adrian Stencel. "A part-dependent account of biological individuality: Why holobionts are individuals and ecosystems simultaneously." *Biological Reviews* 95, no. 5 (2020): 1308-1324.

Swinburne, Richard. *Is There a God?* Oxford: Oxford University Press, 1996.

Taleb, Nassim Nicholas. *Antifragile: Things That Gain from Disorder*. New York: Random House, 2012.

Thomas, Frédéric, A. Schmidt-Rhaesa, Martin Guilhaume Durand Manu, and F. Renaud. "Do hairworms (Nematomorpha) manipulate the water seeking behaviour of their terrestrial hosts?" *Journal of Evolutionary Biology* 15, no. 3 (2002): 356-361.

Tinbergen, Niko. "On Aims and Methods of Ethology," *Zeitschrift für Tierpsychologie*, 20 (1963): 410–433.

Tollefsen, Deborah. *Groups as Agents*. Cambridge: Polity Press, 2015.

Tollon, Fabio. "The artificial view: Toward a non-anthropocentric account of moral patiency." *Ethics and Information Technology* 23, no. 2 (2021): 147-155.

Trivers, Robert L. "The evolution of reciprocal altruism." *The Quarterly review of biology* 46, no. 1 (1971): 35-57.

Turner, Scott. "Individuality in a Social Insect Assemblage." In *From groups to individuals: evolution and emerging individuality*, edited by Frédéric Bouchard and Philippe Huneman, 219-241. Cambridge: MIT Press, 2013.

Velasquez, Manuel G. "Why corporations are not morally responsible for anything they do." *Business & Professional Ethics Journal* 2, no. 3 (1983): 1-18.

Velleman, J. David. "Deciding how to decide." In *Ethics and practical reason*, edited by Garrett Cullity and Berys Gaut. 29-52. Oxford: Oxford University Press, 1997.

------. “How to Share an Intention.” *Philosophy and Phenomenological Research* 57, no. 1 (1997): 29–51.

Waddington, C. H. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. London: Allen & Unwin, 1957.

Walsh, Denis M. *Organisms, agency, and evolution*. Cambridge University Press, 2015.

Ware, Robert. “Review: Collective and Corporate Responsibility.” *The Philosophical Review* Vol. 96, no. 1 (Jan. 1987): 117-119.

Weber, Max. *Max Weber on Law in Economy and Society*. Cambridge: Harvard University Press, 1954.

Wegner, Daniel M. “Transactive memory: A contemporary analysis of the group mind.” In *Theories of group behavior*, edited by Brian Mullen and George R. Goethals, 185–208. New York: Springer New York, 1987.

Well, Nicholas. "Wasp virus turns ladybugs into zombie babysitter." *Science,* February 10, 2015. https://www.science.org/content/article/wasp-virus-turns-ladybugs-zombie-babysitters-rev2.

West, Stuart A., Ashleigh S. Griffin, and Andy Gardner. "Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection." *Journal of evolutionary biology* 20, no. 2 (2007): 415-432.

Wetlesen, Jon. "The moral status of beings who are not persons: A casuistic argument." *Environmental Values* 8, no. 3 (1999): 287-323.

Wheeler, William Morton. "The ant-colony as an organism.” *Journal of Morphology* 22, no. 2 (1911): 307-325.

Wilkinson, Gerald S. "Reciprocal food sharing in the vampire bat." *Nature* 308, no. 5955 (1984): 181-184.

Williams, G. C. *Adaptation and Natural Selection.* Princeton: Princeton University Press, 1966.

Wilson, David Sloan. "Altruism and organism: Disentangling the themes of multilevel selection theory." *The American Naturalist* 150, no. S1 (1997): s122-S134.

------. "Biological communities as functionally organized units." *Ecology* 78, no. 7 (1997): 2018-2024.

------. "Structured demes and the evolution of group-advantageous traits." *The American Naturalist* 111, no. 977 (1977): 157-185.

Wilson, David Sloan, and Elliott Sober. "Reviving the superorganism." *Journal of theoretical Biology* 136, no. 3 (1989): 337-356.

------. "Reintroducing group selection to the human behavioral sciences." *Behavioral and brain sciences* 17, no. 4 (1994): 585-608.

Wilson, Edward O. "Kin Selection as the Key to Altruism: Its Rise and Fall." *Social Research* 72, no. 1 (2005): 159-66.

------. *Sociobiology: The New Synthesis*. Cambridge: Harvard University Press, 1975.

Wilson, Edward O. and Bert Hölldobler. *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies.* New York: W.W. Norton, 2009.

Wilson, Jack. *Biological individuality: the identity and persistence of living entities*. Cambridge University Press, 1999.

------. "Ontological butchery: organism concepts and biological generalizations." *Philosophy of Science* 67 (2000): S301-S311.

Wilson, Robert A. *Genes and the agents of life: the individual in the fragile sciences biology*. Cambridge: Cambridge University Press, 2005.

Wilson, Robert A. and Matthew J. Barker, "Biological Individuals," *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2019/entries/biology-individual.

Wright, Larry. "Functions." *Philosophical Review* 82 (1973): 139-168.

Zilber-Rosenberg, Ilana, and Eugene Rosenberg. "Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution." *FEMS microbiology reviews* 32, no. 5 (2008): 723-735.

Zimmer, Carl. "Meet Nature's Nightmare Mindsuckers." *National Geographic,* November 30, 2014. https://carlzimmer.com/mindsuckers-meet-natures-nightmare-128.