The Philosophy of Feasibility

By

Renan Silva

Submitted to

Central European University

Doctoral School of Political Science, Public Policy and International Relations

Vienna, Austria

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Political Science (Political Theory)

Supervisor: Professor Ferenc Huoranszki

Word Count: 34,596

2023

v. 2

Declaration

I hereby declare that this dissertation contains neither materials accepted for any other degrees in any other institutions nor any materials previously written and/or published by another person, except where appropriate acknowledgment is made in the form of bibliographical reference.

Vienna, 1 October 2023

Renan Silva

Acknowledgments

First and foremost, I would like to thank my supervisor, Ferenc Huoranszki, for his invaluable feedback and discussion and for giving me the freedom to take risks. I am also grateful to the other members of my supervisory panel, Anca Gheaus and Tamás Meszerics, for their very helpful comments. Moreover, I would like to thank everyone who gave me feedback on some of the ideas I had for my dissertation at the very early stages of my PhD, especially Andres Moles, Erin Jenne, Gábor Tóka, János Kis and Zoltan Miklosi. I am also grateful to everyone who commented on my various presentations at CEU, to Hong Do for reading and giving me feedback on a draft of essay three and to the attendees of the 2022 "Ontology of Politics" workshop at the University of Groningen (though the paper I presented there never made it into this dissertation). Finally, I would like to thank my family and friends for their unwavering support.

Abstract

What is it for an outcome to be feasible or infeasible? Some states of affairs, such as world peace, seem to be clearly infeasible, at least in the foreseeable future, while others seem to be evidently feasible, such as my doing my laundry today. Once one moves away from obvious cases like these, things start to get tricky and we need an account of what makes outcomes feasible or not. In this dissertation, I argue that feasibility can be reduced to the agents that exist and the abilities that they have. There is no property of feasibility that exists in its own right – that is, there is nothing extra out there in the world that exists over and above properties such as agency, ability, possibility, likelihood and so on. If one believes in the principle of ontological parsimony, which holds that one should not postulate the existence of more entities than is required, then one should not be committed to a property of feasibility. Based on this finding, I then turn my attention to the nature of agency and abilities. First, I examine one important argument for claiming that organizations (states, firms, political parties, etc.) can be agents and reject it. Then I try to provide an account of one particular kind of ability – state capacity, which has played a significant role in contemporary empirical social science.

Table of contents

Introduction

1. Introduction

Feasibility has figured prominently in various political slogans and movements: from the *Sí, Se Puede* ("Yes, it can be done," in English) slogan of the United Farm Workers Association in Arizona, USA, in 1972, to Barack Obama's *Yes, We Can* slogan for his campaign to the US Senate in 2004 and the US presidency in 2008 (Mettler 2017), to the formation of the *Podemos* ("We can," in English) political party in Spain. There is something about feasibility that captures the popular imagination. If you are interested in a better future, it is not enough that change be desirable. It also needs to be *feasible*. So politicians and campaigners are often keen to emphasize that their policies and movements are within our grasp, that there is something reachable to look forward to if you support their cause. This might have been part of the reason why Obama went along with the slogan *Yes, We Can*, in spite of finding it to be "corny" at first (Mettler 2017).

Feasibility has also captured the imagination of political philosophers in recent years (see Essay 1 for references). They have been particularly interested in the question: does it count against a normative political theory, say a theory of justice or a theory of legitimacy, if it is not feasible? Attempts to answer this kind of question have spawned a large literature and many different sub-debates, one of which being the discussion about the nature of feasibility itself. What is feasibility? The aim of the present dissertation is to contribute towards an answer to this latter question. This area of the feasibility debate has probably been the least developed, if compared to the wider

literature, which has tended to concentrate on the normative/evaluative side of things, instead. Given my focus, I do not claim to be saying anything interesting regarding substantive issues in normative theory.

Nevertheless, I believe that figuring out what feasibility is is philosophically interesting in its own right, and so it is a worthwhile topic of inquiry, independently of normative issues. It is perfectly legitimate to simply want to know whether global justice, open borders, world peace and so on are feasible, regardless of the further question of the desirability of these ends and the moral justifiability of the means to achieving them.

In addition, this dissertation is more oriented towards philosophical concerns. With the exception of essay 3, which provides a chart that outlines a process by which social scientists can select a definition of state capacity, the dissertation does not try to provide clear cut applications to empirical research. Further work would be needed to develop, say, an account of feasibility that is relevant for policy analysis or a definition of group agency that can be operationalized in an empirical model.

In the remainder of this introduction, I will mainly turn to questions of methodology (I will consider a different set of methodological questions in essay 1). Section 2 briefly discusses the history of the word "feasibility," section 3 examines the distinction between philosophical and conceptual analyses, section 4 considers and rejects the use of a philosophical method known as "conceptual engineering" in the context of the debate about the nature of feasibility, section 5 talks about what a good analysis of feasibility should look like and, finally, section 6 provides a short outline of the dissertation.

2. The word "feasibility"

This dissertation investigates the property of feasibility, i.e. the property of *being capable of being done*. This is not the only meaning of the word "feasibility," but it is the most traditional one. The earliest references to the word, initially written in Middle English as *fesable*, *ffeseable*, or *faisible*, come from a document by the Privy Council of England from 1443 and from John Fortescue's 1475 treatise *The Governance of England* (Kurath and Kuhn 1998 [1954], 526). It originated in the Old French word *faisable*, which was formed by adding the present stem of *faire* ("to do," from the Latin *facere*, also meaning "to do") to the suffix *–ble*, a variation upon the Latin *–bilis*, which is a "suffix denoting tendency, fitness, ability, or capability of doing or being something" (Hoad 1986, 43, 168). So the original meaning of "feasible," both in Middle English and Old French was "capable of being done." However, with time, the word acquired a less determinate meaning in English, as the plethora of definitions in contemporary dictionaries of English illustrates.

According to the Merriam-Webster dictionary of English, the word "feasible" has three different meanings: (1) "capable of being done or carried out;" (2) "capable of being used or dealt with successfully: suitable;" (3) "reasonable, likely." The Oxford dictionary defines it as (1) "[p]ossible to do easily or conveniently;" (2) "[l]ikely; probable." The Cambridge dictionary defines it as (1) "able to be made, done, or

achieved;" (2) "possible, reasonable, or likely;" (3) "possible to do and likely to be successful." Similar definitions are also proposed by the Collins dictionary of English.[1]

3. Philosophical analysis versus conceptual analysis

The fact that there are multiple meanings of the word "feasibility" is useful to bear in mind, to avoid any potential confusions, but it will not matter much for this dissertation. As I stated at the beginning, I am concerned with the *property* of being capable of being done. This is called "feasibility" but it could have been called something else. I take it that philosophy is about the analysis of phenomena, as opposed to concepts, words and linguistic expressions (Williamson 2007). Concepts are usually defined as either mental representations or as the meanings of terms (Deutsch 2021a, 11132, 11135). These can be called, respectively, the "psychological" and "platonistic" definitions of concepts (Balaguer and Horgan 2016, 4). As Balaguer and Horgan put it, "[t]he platonistic concept associated with an expression is an abstract entity – in particular, a meaning. […] The psychological concept associated with an expression is a psychological entity that figures as a constituent of intentional mental states" (2016, 4).

Confusingly, *philosophical* analysis is often called "conceptual analysis," even though the former is not about the analysis of concepts but about the analysis of phenomena of philosophical interest (Deutsch 2021a). Many well-known philosophical projects that are usually called "conceptual analysis" are actually instances of

[1] www.merriam-webster.com/dictionary/feasible ; www.lexico.com/definition/feasible ; www.dictionary.cambridge.org/dictionary/english/feasible ; www.collinsdictionary.com/dictionary/english/feasible (Accessed on 27 October 2020).

philosophical analysis, such as Edmund Gettier's attack on the analysis of knowledge as justified true belief, Harry Frankfurt's counterexamples to the argument that moral responsibility entails the possession of free will, the debate over personal identity, the consequentialist analysis of moral rightness (including G. E. Moore's analysis and the various analyses of the trolley problem), Saul Kripke's analysis of proper names and the discussion over whether omissions are causes (Deutsch 2021a, 11128-30). For example, Gettier was not analyzing the *concept* of knowledge – i.e. a mental representation of knowledge or the meaning of the word "knowledge" – but rather knowledge itself (Deutsch 2021a, 11130).

The use of the term "conceptual analysis" to refer to philosophical analysis is so pervasive that some philosophers, like Max Deutsch, suggest that we should just stick to it to avoid further confusion, instead of following the example of Scott Soames (2003a; 2003b), who uses the more accurate term "philosophical analysis" (Deutsch 2021a, 11150n30, 11154n36). However, Deutsch is being too pessimistic. Moreover, even if he is right that throughout most of the history of analytic philosophy, "conceptual analysis" denoted the practice of philosophical analysis, it clearly now also denotes the practice of analyzing concepts (e.g. Jackson 1998). So there would be more confusion in using the same term to refer to two different practices than in using two separate terms. Therefore, it is better to use "philosophical analysis" to refer to the analysis of phenomena of philosophical relevance and "conceptual analysis" to refer to the analysis of concepts.

4. Conceptual engineering

As I stated at the beginning, my project is an instance of philosophical analysis. However, some philosophers might challenge the legitimacy of this kind of inquiry. While some might agree that the analysis of mental representations of feasibility should be left to psychologists and that the analysis of the meaning of "feasibility" should be left to lexicographers and semanticists, one might argue that I should be engaged in the *conceptual engineering* of the concept of feasibility or the word "feasibility" – that is, I should be in the business of improving the word "feasibility" in English (and its analog in other languages) or improving how we mentally represent the phenomenon of feasibility and related phenomena. Herman Cappelen (2018, 3) defines conceptual engineering as the "critical/constructive enterprise of assessing and improving our representational devices," by which he probably means words, linguistic expressions and taxonomies. ("Representational devices" are never defined in his landmark book *Fixing Language*, but he explicitly denies that he is talking about concepts [Cappelen 2018, 3n2, 146-7].) For Cappelen, our talk and thought is plagued by defective "representational devices" and if we are concerned, as he says we should be, about these defects, then conceptual amelioration should be our "most pressing intellectual (and indeed practical) task" (2018, 40) and thus "purely descriptive philosophy must be abandoned" (2018, 47). These defects, he argues, can have deleterious epistemic and normative consequences (Cappelen 2018, 39-43; on the normative argument, see also Haslanger 2000).

The claim that our talk and thought is beset by defective representational devices is an exaggeration (Deutsch 2021a, 11141-2). For example, as Deutsch (2021a, 11142) observes,

> it is implausible to think that *all*, or even very many, of our philosophical concepts are literally meaningless. Perhaps a few are, but not even the positivists held that each and every of our philosophical concepts, and so, presumably, philosophy as a whole, is meaningless. But if most of our philosophical concepts are *not* meaningless, then we ought to be able to use them to think and say true things about what these concepts denote, whatever other defects they may have.

Similarly, just because there is often some degree of vagueness in our terms, it does not mean that we cannot communicate successfully with them (Deutsch 2021a, 11142).

So philosophy as a whole need not be primarily about conceptual engineering, but what about my project? Can a philosophical analysis of feasibility get off the ground without first revising the concept of "feasibility?" Should one even engage in such a philosophical analysis at all, or should one focus entirely on providing the best possible concept of "feasibility?" I will examine the epistemic and normative cases for conceptually engineering the concept of "feasibility," taking each argument in turn.

First, is there something so wrong with the concept of "feasibility" understood as "the property of being capable of being done" that stops us from thinking and saying true things with it? There clearly is not. We can think of actions that can be done (such

as my doing my laundry) and of actions that cannot be done (my running faster than a professional sprinter) and we can successfully communicate about these actions. Perhaps there is an epistemically superior way to think and talk about what can and cannot be done. Maybe one could take each meaning of the word feasibility quoted earlier to be a different concept and then compare them by using Egré and O'Madagain's (2019) general formula for calculating the epistemic utility of concepts/conceptual schemes. One could even add new, made-up (and supposedly superior) concepts to the calculation to see if they are, indeed, superior. However, the burden of proof is on those who wish to show that "the property of being capable of being done" is epistemically inadequate. The only thing I need to demonstrate is that it is not meaningless or utterly confusing.

Second, one could argue that there is a normative case for revising the concept of "feasibility." On the one hand, one could maintain that an adequate concept of "feasibility" should not rule out some otherwise desirable political project as infeasible. If it does, then there must be something morally wrong with the concept. For example, the most popular criticism of socialism is that it is infeasible. Indeed, in popular debates, the infeasibility argument against socialism is often made as a "gotcha!" moment intended to settle the debate for good. So a socialist might argue that the concept of "feasibility" should be revised to circumvent these kinds of challenges. On the other hand, one could contend that the concept of "feasibility" is too broad and allows many undesirable projects to be judged as feasible, such as the creation of a global caliphate. If they were *not* considered to be feasible and were known not to be, then that might dampen people's motivation to pursue these nefarious projects. So, again, perhaps there

is a normative case to revise the concept of "feasibility" to make it rule out morally wrong political aims.

It is important to distinguish here between conceptual engineering in politics and activism and in academia. Politicians and activists often change (or try to change) the meaning of words to suit their political agendas. Sometimes this is morally right, but it is usually not. For example, it is common for politicians to change the official definition of terms like "unemployment" and "crime rate." It will depend on the context whether it is morally right for them to do that. Perhaps a false or "revised" portrayal of how the economy is performing is necessary in the extreme scenario in which one needs to prevent the electoral success of a fascist party, for example. By contrast, academics should concern themselves with the impartial pursuit of truth and knowledge (van der Vossen 2015; 2020; but see Jones 2020 for the opposite view). Conceptual engineering in academic research for normative reasons is a form of what Jon Elster called "soft obscurantism" (2015, 454-8). While "hard obscurantism" is characterized by practices such as the atheoretical misuse of mathematical tools in rational choice theory, "soft obscurantism" is characterized by an academic disregard for the norms for arriving at truth or simply by a lack of interest in the pursuit of truth (Elster 2015, 452-91). In this latter, more extreme variety, soft obscurantism adheres to Humpty-Dumptyism:

> [Those who scorn the very idea that there is such a thing as truth] would endorse the response of Humpty Dumpty to Alice when she said, "the question is whether you can make words mean so many different things." "The question is," he answered, "which

> is to be the master; that's all." *Power*, not truth, determines which theories will succeed (Elster 2015, 454).

Some advocates of conceptual engineering like Cappelen argue that the charge of Humpty-Dumptyism can be avoided if a conceptual revision only changes the meaning and extension of a concept or word but does not change its *topic* (Cappelen 2018, ch. 10). He argues that only if one ignores topic continuity does one risk the spread of deception or a breakdown in communication (Cappelen 2018, 132-4). I believe that he is right that in *some* cases we can say the same thing when using a term X even though we do not ascribe the same meaning and extension to X. The case of "salad" is a compelling one: two people can both say the same thing when uttering that "salad is delicious," even though one person is referring to cold dishes made up primarily of green leaves while another is referring to a broader category that also includes dishes that are warm and do not include leaves (Cappelen 2018, 112-3, citing Dorr and Hawthorne 2014). Think of how, in the past, "salad" had a more restricted meaning and extension, but someone in 2023 could still quote truthfully a speaker from the 1950s saying that "salad is delicious" even though the 2023 speaker has a broader understanding of "salad." Cappelen generalizes the argument as follows (2018, 108):

> "extension," "intension," and "content" are *theoretical* terms, and are not things on which we have a pre-theoretic grasp. By contrast, expressions like "what she said," "what she was talking about," and "talking about the same topic" are important pre-

> theoretic notions. These *pre-theoretic* notions are more coarse-grained than the theoretical notions that philosophers have used in the last one hundred years. "What was said by utterance *u*" is much more coarse-grained than "the semantic content of *u*" (relative to context). As a corollary, two sentences with different semantic contents (where semantic content is understood as, at least, having the same extension and intension) can be used to say the same thing, or to talk or be about the same topic.

My problem with this is that although it is true in some cases, in many other cases it is not, including the case of the concept of "feasibility." Suppose a university hiring committee is looking for a historian specializing in British history since 1945. A potential job candidate who specializes in Anglo-Saxon England could then apply to the job and claim that he is an expert in the topic of the job advert – British history. But it is obvious that what matters here is the meaning and extension of "British history since 1945" and not sameness of topic, even though the job candidate could, perhaps, truthfully say "I applied for a job in my area of expertise" and by "area of expertise" mean "the history of Anglo-Saxon England." Similarly, someone who uses the term "feasibility" to mean "likelihood of success" and another who uses it to mean "ability of being done" can, perhaps, be talking about the same thing when each says "Feasibility is a constraint on a theory of justice." For example, they could both be saying that there is something far-fetched about a theory of justice that requires the end of private property, the complete abolition of gender roles, the end of racism and the

achievement of equality of outcomes. That is, they could both be talking about the same topic when uttering that "Feasibility is a constraint on a theory of justice," but meaning different things by the word "feasibility." However, this sameness of topic is not enough for rigorous research in philosophy and science. Suppose one says that it is feasible to turn Finland into a libertarian minimal state in the next five years. Is it? If "feasibility" is "likelihood of success," then it clearly is not feasible. By contrast, if "feasibility" is "the ability of being done" then it is not straightforward what the answer is and more empirical research would be needed. Scholarly research needs sameness of *semantic* content, not sameness of topic. To insist on sameness of topic and to do so for political reasons is an exercise in soft obscurantism.

Moreover, advocates of conceptual engineering as a philosophical methodology should be happy to embrace a concept of "feasibility" with a settled meaning and extension. After all, a key challenge to this project is to show that it is feasible (Deutsch 2020; 2021b; Fischer, forthcoming; Gibbons, forthcoming; Jorem 2021; Koch 2021a; 2021b; Koslow 2022; Machery, forthcoming; Nimtz, forthcoming; Thomasson, forthcoming; but see Andow 2021; Cappelen 2018, 72-8; Pinder 2021; Queloz and Bieber 2022; Riggs, forthcoming). That is, that changes in meaning can be intentionally brought about, either by the conceptual engineers themselves or by others. If defenders of conceptual engineering attempt to determine whether conceptual engineering is feasible or not by first *revising* the concept of "feasibility," they will beg the question. However, if "feasibility" has a settled meaning as "being capable of being done," then at least that problem can be avoided for their project. Nevertheless, some philosophers are happy to take their argument to its logical conclusion and simply beg the question.

Cappelen, for example, writes that "all the concepts involved in describing the critical/constructive project of conceptual engineering should themselves be subject to constant critical assessment and skepticism. […] The very terminology in which you engage in the critical project is itself suspect" (2018, 48). My impression, though, is that at least some advocates of conceptual engineering would prefer not to pay such a high price for their philosophical commitment. Indeed, the existing literature on the feasibility of conceptual engineering is relatively silent on whether "feasibility" should be revised or not.

5. What does a successful analysis of feasibility look like?

So an analysis of feasibility is neither an analysis of the concept of "feasibility" nor is it a project on the amelioration of that concept. Rather, it is about the property of feasibility itself.

There is a common misconception that for a philosophical analysis to be successful, it must be a *full* analysis – an analysis of all the necessary and sufficient conditions of a phenomenon. This assumption is unjustified because it is perfectly legitimate to provide only one significant necessary condition or one relevant sufficient condition, or to show the absence of such conditions, as Gettier did when he challenged the sufficiency of justified true belief for knowledge (Deutsch 2021a, 11133, 11145-7).

My dissertation is only a partial analysis of feasibility. It argues that feasibility can be reduced to agency and abilities (essay 1), but it only focuses on the agency of organizations such as states, firms and political parties (essay 2) and on the abilities of

states (essay 3). A *full* analysis would have to say much more about the nature of individual and collective agency and the nature of individual and collective abilities. For example, in essay 2, I criticize one argument for the view that organizations can be agents, but I do not actually take a stand on whether other arguments for organizational agency succeed or fail. In essay 3, I come out in favor of one existing analysis of individual abilities and extend it to the case of states, but I do not provide a fully-fledged defense of it in light of the wider literature on individual abilities. A truly complete analysis of feasibility would require many thousands of pages' worth of arguments, given that it would need to draw upon very large and complex bodies of literature, from metaphysics and the philosophy of agency to the philosophy of social science. So it is no surprise to say that I will not be doing that here and will be sticking to my more modest aims.

6. The structure of the thesis

Although this dissertation is fairly cohesive in its subject matter, it is written as a collection of essays that can be read independently. Each essay contains its own abstract, so I will provide only a brief summary here. The first essay investigates whether we should include in our ontology a property of feasibility. It argues that we should not because feasibility can be reduced to the agents that exist in the actual world and their abilities. It contends that most of the philosophically interesting questions about feasibility are really questions about agency and abilities. With that in mind, I then go on, in the second essay, to focus on one kind of agency, the agency allegedly

had by organizations, such as states, firms and political parties. I examine one argument for accepting that organizations can be agents, which draws upon a view in the philosophy of science known as "scientific realism" and upon examples from empirical social science that some scholars believe vindicate realism about organizational agency. I find that the argument fails but I also provide a route for rehabilitating it. Then, in the third essay, I develop an account of the abilities of states, which builds upon an analysis of individual abilities known as the "success view" and which takes into consideration the different understandings of the nature of the modern state, from those that see the state as a non-agential system to those that see it as an intentional agent. Following that article, I provide a conclusion with some final reflections about the whole dissertation.

References

Andow, James. 2021. "Conceptual Engineering Is Extremely Unlikely to Work. So What?" *Inquiry* 64 (1-2): 212-26.

Balaguer, Mark and Terry Horgan. 2016. "A Solution to the Paradox of Analysis." *Analysis* 76 (1): 3-7.

Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.

Deutsch, Max. 2020. "Speaker's Reference, Stipulation, and a Dilemma for Conceptual Engineers." *Philosophical Studies* 177 (12): 3935–57.

Deutsch, Max. 2021a. "Conceptual Analysis without Concepts." *Synthese* 198 (11): 11125–57.

Deutsch, Max. 2021b. “Still the Same Dilemma for Conceptual Engineers: Reply to Koch.” *Philosophical Studies* 178 (11): 3659–70.

Dorr, Cian and John Hawthorne. 2014. “Semantic Plasticity and Speech Reports.” *Philosophical Review* 123 (3): 281–338.

Egré, Paul and Cathal O’Madagain. 2019. “Concept Utility.” *Journal of Philosophy* 116 (10): 525-54.

Elster, Jon. 2015. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Revised Edition. Cambridge: Cambridge University Press.

Fischer, Eugen. Forthcoming. “Conceptual Control: On the Feasibility of Conceptual Engineering.” *Inquiry*. doi.org/10.1080/0020174X.2020.1773309

Gibbons, Adam. Forthcoming. “Rational Conceptual Conflict and the Implementation Problem.” *Inquiry*. doi.org/10.1080/0020174X.2022.2037455

Haslanger, Sally. 2000. “Gender and Race: (What) Are They? (What) Do We Want Them to Be?” *Noûs* 34 (1): 31-55.

Hoad, T. F. 1986. *The Concise Oxford Dictionary of English Etymology*. Oxford: Oxford University Press.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

Jones, Ben. 2020. “Political Activism and Research Ethics.” *Journal of Applied Philosophy* 37 (2): 233–48.

Jorem, Sigurd. 2021. “Conceptual Engineering and the Implementation Problem.” *Inquiry* 64 (1-2): 186-211.

Koch, Steffen. 2021a. “The Externalist Challenge to Conceptual Engineering.” *Synthese* 198 (1): 327–48.

Koch, Steffen. 2021b. “There Is No Dilemma for Conceptual Engineering. Reply to Max Deutsch.” *Philosophical Studies* 178 (7): 2279–91.

Koslow, Allison. 2022. “Meaning Change and Changing Meaning.” *Synthese* 200, Article Number: 94.

Kurath, Hans and Sherman Kuhn. 1998 [1954]. *Middle English Dictionary*. Ann Arbor: The University of Michigan Press.

Machery, Edouard. Forthcoming. “A New Challenge to Conceptual Engineering.” *Inquiry*. doi.org/10.1080/0020174X.2021.1967190

Mettler, Katie. 2017. “Obama’s ‘Yes We Can’ almost didn’t happen. You can thank Michelle for saving it.” Available at: https://www.washingtonpost.com/news/morning-mix/wp/2017/01/11/obamas-yes-we-can-thank-michelle-for-that/ (accessed on 27 October 2020).

Nimtz, Christian. Forthcoming. “Engineering Concepts by Engineering Social Norms: Solving the Implementation Challenge.” *Inquiry*. doi.org/10.1080/0020174X.2021.1956368

Pinder, Mark. 2021. “Conceptual Engineering, Metasemantic Externalism and Speaker-Meaning.” *Mind* 130 (517): 141–63.

Queloz, Matthieu and Friedemann Bieber. 2022. “Conceptual Engineering and the Politics of Implementation.” *Pacific Philosophical Quarterly* 103 (3): 670-91.

Riggs, Jared. Forthcoming. “Conceptual Engineers Shouldn’t Worry about Semantic Externalism.” *Inquiry*. doi.org/10.1080/0020174X.2019.1675534

Soames, Scott. 2003a. *Philosophical Analysis in the Twentieth Century: Volume 1: The Dawn of Analysis*. Princeton and Oxford: Princeton University Press.

Soames, Scott. 2003b. *Philosophical Analysis in the Twentieth Century: Volume 2: The Age of Meaning*. Princeton and Oxford: Princeton University Press.

Thomasson, Amie. Forthcoming. “Conceptual Engineering: When Do We Need It? How Can We Do It?” *Inquiry*. doi.org/10.1080/0020174X.2021.2000118

van der Vossen, Bas. 2015. “In Defense of the Ivory Tower: Why Philosophers Should Stay Out of Politics.” *Philosophical Psychology* 28 (7): 1045-63.

van der Vossen, Bas. 2020. “Academic Activism Revisited.” *Journal of Applied Philosophy* 37 (2): 249-57.

Williamson, Timothy. 2007. *The Philosophy of Philosophy.* Oxford: Blackwell Publishing.

Essay 1: Is There a Property of Feasibility?

Abstract

In ordinary language, feasibility, which can be defined as the property of being capable of being done, is often ascribed to states of affairs and actions. But is there really a property of feasibility that exists in its own right? In this article, I argue that there is not. If one accepts a prominent version of actualism and if one is committed to the existence of agency and abilities, then one has little reason to include a property of feasibility in one's ontology. In the process of arguing for this claim, I also try to bring greater clarity to the debate over the nature of feasibility by suggesting how feasibility claims should be interpreted. I note, for example, that while feasibility claims entail ability claims, they are not logically equivalent. Moreover, I address questions of methodology and criticize the method of trying to settle disputes about the nature of feasibility by examining the function that our talk and thought about feasibility plays.

1. Introduction

We often talk about certain outcomes being feasible or infeasible. From mundane outcomes such as an increase in the productivity of a firm to more radical ones such as socialism and open borders, we often ascribe the property of feasibility or infeasibility to them. While there has been growing interest among philosophers in recent years in

the nature of feasibility,[1] little attention has been paid to the question of whether feasibility is a distinctive thing that exists in its own right. That is, is there a *sui generis* property of feasibility? Or is talk of feasibility merely a manner of speaking, a way of paraphrasing claims about agents' abilities and dispositions or claims about the necessity, possibility, contingency or probability of certain states of affairs? In this article, I will argue that there is no property of feasibility in addition to the more common modal properties postulated by metaphysicians, such as ability and possibility.

First, I outline a general metaphysical framework against which claims about the existence of the property of feasibility should be judged (section 2); then, I examine what an alleged property of feasibility could amount to, once one deflates merely verbal disputes about feasibility (section 3); in section 4, I examine an alternative methodology for settling debates about the nature of feasibility, functionalism, and reject it as unpersuasive; then section 5 argues, on general grounds, that there is no *sui generis* property of feasibility; finally, section 6 contends that even if one accepts that the recent

[1] Some representative works include Brennan and Sayre-McCord (2016); Erman and Möller (2020: 3-6); Estlund (2020: 243-8); Gheaus (2013: 449-51); Gilabert and Lawford-Smith (2012: 810-8); Guillery (2021); Hamlin (2017); Lawford-Smith (2012: 463-5; 2013; 2022: ch. 8); Räikkä (1998); Silva (2019); Southwood and Wiens (2016); Southwood (2016: 11-7; 2022); Stemplowska (2016; 2021); Wiens (2015). On the distinct, but related, issue of feasibility's role in political philosophy, see Southwood (2018) and Erman and Möller (2020) for good overviews and further references. Also note that many philosophers have discussed the nature of feasibility, but under a different terminology. For example, while Estlund (2020: 243-8) defines "feasibility" as the ability of groups that are not agents and then proceeds to provide an account of these non-agentive group abilities, Collins (2019: 71; forthcoming: §2.2) also discusses non-agentive group abilities but does not use the language of "feasibility."

account of feasibility proposed by Southwood (2022) is true, there is no reason to suppose that it lends support to the thesis that there is a property of feasibility that exists in its own right.

2. Background: actualism, states of affairs, events and actions

The question of whether there is a property of feasibility or not can only be answered if one takes a stand on certain controversial metaphysical disputes. For example, since the beginnings of philosophy, there has been discussion about the existence of things such as properties and relations. Many philosophers deny their existence and argue that only concrete objects (also known as "particulars") such as tables or apples exist, while others defend the existence of properties, relations and other abstract entities (see e.g. Falguera, Martínez-Vidal and Rosen 2022; MacBride 2020; Orilia and Paoletti 2022). Here I will side with the latter.

More specifically, I will take for granted a view that is known as *actualism*, which is relatively popular among metaphysicians. However, there are many different versions of it and little agreement as to which one is the best. For present purposes, the specifics will not matter that much but I will focus on a specific account for the sake of simplicity, namely that defended by Alvin Plantinga (1974; 1976), which has been quite influential in the literature. According to it, we should take at face value the existence of things such as truth, properties, relations, propositions, states of affairs, possibility, necessity and contingency and we can illuminate the nature of these things by studying each of them in terms of the other, even though neither can be *reduced* to the other.

Moreover, we should accept the existence of abstract objects in addition to concrete ones. Abstract objects include things like the number 4 and Mozart's symphonies. They also include things like properties (e.g. the property of being green) and states of affairs (e.g. *Whales' being animals*). One should distinguish, however, between (a) an abstract object's existing and (b) its being exemplified/instantiated or its obtaining. The property of being green exists but it is only instantiated if there is actually an object with that color. Similarly, the state of affairs *Mitt Romney's winning the 2012 US presidential election* exists but it never obtained and it never will. A *world* is the totality of states of affairs that exist. A *possible* world is the totality of states of affairs that exist and that are possible to obtain alongside one another. The *actual* world is the totality of states of affairs that exist and obtain. Our world is the actual world. The state of affairs *Jeff Bezos' giving me one billion dollars* is possible if there is some world that, were it to obtain, Jeff Bezos would give me one billion dollars. It is impossible if there is no world in which Bezos gives me that money, even if that world obtains. These are the key aspects of Plantinga's actualism that are worth noting for current purposes.

However, one important amendment should be made to it. We should abandon Plantinga's assumption that *all* states of affairs necessarily exist, regardless of whether the objects and properties that they involve also exist. It is more plausible to say that a state of affairs exists only if the objects that it involves also exist (Fine 1976: 563).[2] But if we discard the notion that all states of affairs necessarily exist, we need an account of

[2] Note that Fine uses the term "proposition" to refer to state of affairs (Textor 2021: §5.3).

what makes some combination of objects and properties give rise to a state of affairs while others do not.

One compelling strategy here is to appeal to the notions of *ontological dependence* and *predicability* (Fine 1982: 51-2; Textor 2021: §5.3, §6.3). A simple (or "atomic") state of affairs such as *the cat's being black* is an entity that exists in virtue of the cat (an object) and the property of being black. It ontologically depends on the cat and the property of being black. But it is not a whole that has the cat and the property of being black as its parts, so we do not run the risk of being committed to the idea that the state *the cat's being black* (an abstract object) has the same parts as the cat (a concrete thing), such as whiskers, which would be absurd. So the relation between the elements of a state of affairs is one of ontological dependence, as opposed to one of constitution (or some other relation) (Fine 1982: 51-2; Textor 2021: §5.3). And what determines *which* states of affairs exist is the nature of the objects and the properties in question (Fine 1982: 51-2; Textor 2021: §6.3). *The cat's being black* exists because it is part of the nature of cats and of the color black that cats can be black. In the jargon, we say that the property of being black is *predicable* of a cat and that gives rise to the state of affairs *the cat's being black*. The appeal to predicability avoids counter-intuitive alternatives to characterizing the existence conditions of states of affairs, such as that mental acts or language can unify an object and a property in a state of affairs (Textor 2021: §6.2).

What about complex (or "molecular") states of affairs, which are states of affairs that involve other states of affairs, such as *the Allies' winning the war*? The identity and existence of these states of affairs "are the result of the application of such operations

as conjunction and disjunction to basic states of affairs" (Textor 2021: 5.3; cf. Fine 2010: 586-8). For example, *the Allies' winning the war* just is *the Red Army's taking over Berlin* and *soldier X firing artillery in location Y* and *artillery shells' being explosive*, and so on. Note that this is true even if *the Allies' winning the war* had never obtained and even if it had been impossible to obtain.

What about events and actions? An *event* is the change that takes place when an obtaining states of affairs ceases to obtain because an object no longer exemplifies the property that gave rise to it or when a non-obtaining state of affairs obtains (cf. Ludwig 2016: ch. 2). So *the Allies' winning the war* became the event of the allied victory in World War II once that state of affairs changed from not obtaining to obtaining in 1945. An *action* is a kind of event, namely one that is intentionally brought about by an agent (Davidson 2001 [1967]; but see Alvarez and Hyman 1998 for criticism). While the allied victory is an event, it is not an action because it cannot be attributed to an agent. However, many of the events that the allied victory involve were actions, such as the firing of artillery by a particular soldier at a given time in Berlin. One might argue that the allied victory *is* an action because "the Allies" should be understood as a group agent in its own right, an entity that exists over and above the individual soldiers, politicians, citizens and so on that constituted the Allies. For the purposes of this article, though, this issue does not matter. We can grant that there are such things as group agents. What is key is that actions are a kind of event that is attributable to an agent but that there are events that are not brought about by an agent.

Finally, it is important to clarify that while objects, human beings, non-human animals and other concrete particulars are the bearers of modal properties such as

abilities and dispositions, states of affairs, events and actions are the bearers of modal properties such as possibility, contingency, necessity and probability (see Textor 2021: §2.2). A human being can do things and a glass sculpture can be disposed to break, but a state of affairs cannot be the bearer of abilities and dispositions. On the other hand, a human being cannot be possible or impossible, contingent or necessary, only states of affairs, events and actions can.

Given this fairly permissive ontology, one wonders what role feasibility could play, if any, in this framework. In the next section, I will argue that there is no need to postulate a property of feasibility. An ontology that includes states of affairs, the traditional modal properties and agents is more than enough to play whatever role is suggested by talk of feasibility. To put it differently, feasibility can be explained away and so there is little reason to accept the existence of a property of feasibility - at least, if one is committed to the widely held principle of ontological parsimony, which is the idea that one should not postulate more entities than is explanatorily required (for example, one need not resort to the notion that there are evil spirits in order to explain the occurrence of an epidemic, so one should discard such evil spirits from one's ontology).

3. The property of feasibility

If there were a property of feasibility, it would be borne by things like states of affairs and events (including actions), as opposed to human beings and objects. For example, socialism is feasible or infeasible and a person's action is feasible or infeasible.

Socialism here is understood as a non-obtaining state of affairs that is the conjunction of other non-obtaining states of affairs, such as the abolition of private property, the allocation of resources based on needs, the production of resources based on ability and so on. A person's action is a change in states of affairs so it is this change that would be the bearer of the property of feasibility and not the person herself.

In English, people mean different things by the word "feasibility" (and "feasible"). For example, the predicate "feasible" in the sentence "Socialism is feasible" has at least the following meanings:

(1) Socialism is likely.

(2) Socialism is highly likely.

(3) Socialism is possible.

(4) Socialism can be intentionally brought about.

(5) Socialism can easily be intentionally brought about.

(6) Socialism is highly likely and desirable.

(7) Socialism is desirable and neither morally costly nor risky.

Given that in the ontology taken for granted in this article, we already accept the existence of traditional modal properties such as possibilities and objective probabilities, the word "feasible" does not pick out anything distinctive that is of philosophical significance. Understood as a modal property, it merely denotes possibility or likelihood, so any discussion about the nature of feasibility amounts to a verbal dispute that can be easily deflated if we avoid using the word "feasible" altogether (see Chalmers 2011). The same goes for the interpretation of "feasible" as "highly likely and desirable": a state of affairs' being desirable and highly likely to

obtain can be analyzed in terms of familiar modal properties (e.g. probabilities) and normative/evaluative properties (e.g. desirability).

What *would* be distinctive and philosophically significant would be if there were a property of being capable of being intentionally brought about – that is, a property of *being capable of being done (or accomplished, carried out, realized, etc.)*. This is quite different from possibility and likelihood and has been studied much less extensively by philosophers. So by "feasibility" I will mean this property of being capable of being done, but one can use whatever word one wishes to capture this property.

4. The case against functionalism about feasibility

Some philosophers are suspicious of attempts to disambiguate between the different meanings of the word "feasibility" to check whether there is more at stake than a merely verbal dispute. Southwood (2022: 161-2) claims that this strategy is an example of the discredited linguistic approach to philosophy and that we cannot shed light on the nature of feasibility by looking at the meaning of words. However, his position is not compelling. Looking at the meaning of the word "feasibility" in English and selecting one of its meanings as the most important one for philosophical analysis is just a way of delimiting the scope of inquiry. It is just like choosing to limit the scope of an investigation into the history of European "rulers" to *political leaders*, instead of including both political leaders *and* the physical tool used to measure the length of objects. One is not engaging in semantic analysis here, just delimiting what one's inquiry is going to be about.

Someone might insist, though, that "feasibility" is not like "ruler" because the former, unlike the latter, has meanings that are somewhat related – they all have something to do with modality. So disambiguating between the different meanings of "feasibility" is an illegitimate way of trying to shed light on the nature of feasibility. It is arbitrary and begs the question in favor of one's preferred interpretation of feasibility (the property/phenomenon). The fact that the word "feasibility" denotes (among other things) both likelihood and capability of being done is irrelevant to settling the question of what feasibility truly is. Perhaps it *should* be identified with both likelihood *and* being capable of being done. Starting with a "disambiguation" between the different meanings of the word "feasibility," critics might say, might end up obscuring the fact that there is nothing to disambiguate, that feasibility *is* an emergent property that depends upon or is constituted by the property of being likely, the property of being capable of being done, the property of being likely and desirable, etc.

I agree that someone might be able to provide an argument for the existence of this emergent property. Nevertheless, it remains the case that the word "feasibility" has different meanings and that it is a legitimate form of inquiry to focus on just one of them. More specifically, it is legitimate to probe into whether one alleged entity that the word "feasibility" picks out, namely *the property of being capable of being done* is a property that exists in its own right or not. Whether there is a further, emergent property that depends upon or is constituted by it is a separate question. There is nothing suspect about disambiguating between the different meanings of "feasibility" and doing so does not amount to doing linguistic philosophy. After all, a parallel argument could be made that disambiguating between the different meanings of the word "ruler" and focusing

on just ruler-as-political-leader could obscure the fact that there might be an emergent object that is made up of Napoleon and his golden ruler (just as there might be an object constituted by my face and my microwave). For it might turn out that universalism, the view that "for any objects, there is a single object that is composed of those objects" (Korman 2020: §1.3), is true.

Southwood (2022) argues that instead of starting by disambiguating between the different meanings of the word "feasibility," we should adopt a functionalist approach to how we theorize about feasibility. We can evaluate different philosophical accounts of what feasibility is, he maintains, by examining what central functional role our talk and thought about feasibility plays. That is, the way we talk and think about feasibility has a key purpose. So a good account of feasibility is one that accommodates this purpose and we can rule out those that do not. He contends that the functional role of our talk and thought about feasibility is to determine which actions are worthy of deliberation. For example, the claim or thought that it is not feasible for me to jump from a 10-meter diving platform plays the role of ruling out that action as something worthy of deliberation for me. Jumping from the platform is something that is beyond my "domain of deliberative jurisdiction." By contrast, the claim or thought that eating a tub of ice cream (when there is ice cream in my freezer) is feasible for me plays the role of including that action as an option that is worthy of deliberation for me.

Southwood maintains that we do not need to settle what the predicate "feasible" means – whether it means possible, likely, desirable, not morally costly or whatever – to determine if an action is deliberation-worthy and, therefore, *feasible*. For feasibility just is a *functional* property that can be realized in a number of different ways. Its

functional role – deliberation-worthiness – can be realized by various modal properties.[3] For example, it might be the case that it is realized by a restricted possibility in one context but by a disposition in another. The infeasibility of my jumping from a 10-meter diving platform could, perhaps, be said to be realized by my disposition to be afraid of heights. But the feasibility of my eating ice cream could, perhaps, be said to be realized by a restricted possibility, namely the possibility of eating ice cream when there is ice cream around. All that matters is whether the action is deliberation-worthy or not, as opposed to how deliberation-worthiness manifests.

Southwood goes on to argue that existing accounts of feasibility, which equate feasibility with a specific property, such as possibility or likelihood, fail because they either include or exclude too many actions from the domain of the deliberation-worthy. For example, a simple probability account of feasibility, according to which an action is feasible for me only if I am likely to perform it, objectionably rules out many actions that are perfectly deliberation-worthy. For instance, I am unlikely to wake up at 5 am every day to run 10km but that is still deliberation-worthy. Meanwhile, a restricted possibility account takes far too many actions to be feasible, given that only performance in a possible world is required, even though these actions are not worthy

[3] Although some definitions of "feasibility" also include evaluative/normative properties such as desirability and low moral costliness, Southwood suggests that evaluative/normative properties do not realize the function of deliberation-worthiness/feasibility, because they are too exclusive. For example, leaving one's petulant toddler alone at the supermarket is deliberation-worthy/feasible but morally beyond the pale (Southwood 2022: 138-9). Only modal properties such as possibility and likelihood can realize the role of deliberation-worthiness/feasibility (Southwood 2022: 150).

of deliberation. For example, there is a possible world in which I model for Calvin Klein full-time and make millions but that is hardly deliberation-worthy.

I will leave aside the substantive question of whether there is a property of deliberation-worthiness that is borne by some actions but not by others. I want to focus, instead, on Southwood's argument about methodology. He is definitely right that his functionalist approach is an improvement on the method of coming up with sentences with the predicate "is feasible," then trying to judge whether those sentences sound felicitous or infelicitous based on how we feel about them and then trying to derive some major philosophical conclusion from these feelings. However, Southwood's functionalism is also an objectionable kind of linguistic approach to philosophy. First, *whose* talk and thought of feasibility is he referring to? It seems that he is implicitly making the universal claim that human beings talk and think about feasibility in a certain way. But this is an empirical claim for which he provides no evidence. Some languages, such as Portuguese and Spanish, do not even have a direct translation for the word "feasible." The word is usually translated as "viable" (*viável* in Portuguese and *viable* in Spanish), which is not quite the same as "feasible." For example, a central meaning of "viable" is the ability to carry on living or to become a living being (e.g. "the embryo is viable").

More plausibly, Southwood's claim about "our talk and thought about feasibility" is not a claim about human behavior in general, but about specific linguistic communities, such as the community of English speakers (or perhaps, the community of English-speaking political philosophers). He writes that "unlike the linguistic approach, [functionalism] involves attending to our practices of using claims about

feasibility, whatever their content, to do things (what I shall call our practices of *practical deployment*)" (Southwood 2022: 124). The most natural way of interpreting this claim is that he is interested not in semantics but in pragmatics. But why should the pragmatics of English or any other language have any bearing on debates about what feasibility is? It seems odd that the way an English speaker in Lagos or Sydney uses feasibility statements – whether they use them to determine the deliberation-worthiness of an action, to undermine the policy proposals of one's political opponents or some other purpose – should be the standard against which accounts of feasibility should be judged.

Consider an analogy. Suppose we applied this functionalist approach to debates about what the color red is. We would start by ignoring the meaning and extension of the word "red" in English so that we could avoid taking a stand on controversies about settling the point at which red becomes brown, pink or orange. For example, we could ignore whether the color Burgundy is picked out by the word "red" or "brown." Instead, we would try to identify a central function that English speakers' talk and thought of red plays. (Or imagine that we can be extremely thorough and identify the functional roles of the talk and thought of *all* linguistic communities that ever existed, including the Ancient Greeks, who famously had a different understanding of color to that of English speakers.) Suppose that we identify three key functional roles: the role of distinguishing the poisonous red berries from the edible blue berries; the role of determining when it is safe to move one's motorized vehicle at a traffic light; and the role of spotting blood.

Now, on Southwood's approach, a good account of what the color red is should be able to accommodate these three fundamental functional roles. However, this implies that what the color red is is somehow subjective and relative to the practices of linguistic communities. For the persuasiveness of an account of red would change depending on how linguistic practices change. This is an unsatisfactory conclusion for those who are looking for an objective and universal account of the property of being red.

Note that Southwoodian functionalism about color is distinct from one of the most prominent contemporary functionalist analyses of color, proposed by Brian McLaughlin (2003). McLaughlin argues that redness is a property of objects that has a certain function in visual experience. His argument is not based on the notion that "our talk and thought" about redness plays a certain functional role, but on the idea that redness *itself* has a particular function. He describes his view as follows:

> Redness is a visual property in that it plays a certain role vis-à-vis visual consciousness: namely, the role of being the property that disposes its bearers to look red to standard visual perceivers in standard conditions of visual observation and that (nomologically) must be possessed by everything so disposed. Call this role 'the redness-role'. According to the basic proposal, then, redness is just that property, whatever it is, that occupies the redness-role. (McLaughlin 2003, 479)

This means that McLaughlin's functionalism is not subject to my criticism regarding linguistic relativism because it does not rely on time-dependent and contingent linguistic practices.[4]

When it comes to feasibility, linguistic relativism is particularly unsatisfactory. There is something extremely underwhelming about a claim such as "a world of open borders is (not) feasible, but only given the current behavior of existing linguistic communities." But what linguistic communities do or not seems neither here nor there. Something is either feasible or infeasible, regardless of linguistic behavior.

Moreover, there is something objectionably anthropocentric about Southwood's functionalist methodology. It is now a widely held view that sophisticated non-human animals can perform intentional actions or something very close to them (Piñeros Glasscock and Tenenbaum 2023: §8). If that is true, then it makes sense to make feasibility statements about non-human animals. For example, we can say that it is or it is not feasible for a specific pack of wolves to catch a deer in the forest or for a fish to escape an orca before it is captured. Again, it is strange to claim that how an English speaker in Lagos or Sydney practically deploys the word "feasible" can help illuminate whether the fish can escape from the orca or not. Surely, even if all human beings were

[4] To be sure, one could argue that McLaughlin's account is relativist in the sense that it depends on contingent facts about the nature of biological organisms and other systems capable of visual experience. However, if this is true, it is a kind of relativism that one can live with. The biological fact that human beings cannot run at 500km/h is contingent and could have been false if evolution had panned out differently. But it is a very stable fact, unlike facts about linguistic practices, which are extremely fluid and time-dependent.

wiped off the face of the Earth, there would still be a fact of the matter about the feasibility of the fish's escape that is language- and mind-independent.

I am not saying that there can be *no* functionalist approach to feasibility. Perhaps a functionalist view could be proposed according to which feasibility is identified with a certain function but that function has nothing to do with the practices of linguistic communities. Maybe one could try to develop a functionalist view of feasibility along the lines of McLaughlin's functionalism about color. But it remains to be seen whether anyone can provide a compelling version of this view.

5. Why there is no property of feasibility

Having argued for the respectability of my approach to how we should theorize about feasibility and rejected one prominent alternative, we can now turn to the substantive question of whether there is a property of feasibility.

Consider the state of affairs *Socialism's being feasible*. It can be interpreted in one of two ways:

> (1) There is an individual agent, collection of individual agents or group agent *in the actual world* that can bring about socialism.
>
> (2) There is an individual agent, collection of individual agents or group agent *in some possible world* that can bring about socialism.

The second interpretation is not very plausible. For there are possible worlds in which a group of socialist activists are trillionaires and politically powerful and, therefore, can single-handedly accomplish socialist policies, but this is not particularly illuminating. It does not seem to tell us anything substantively interesting about whether socialism is or not feasible. It merely states a counterfactual claim, namely that had a non-obtaining world in which there are extremely wealthy and powerful socialist activists obtained (i.e. had this world become the actual world), then socialism would also have obtained.

By contrast, the first interpretation provides more than a counterfactual claim about socialism – it tells us how things really are in our world. If that is true, then the next step is to try to elaborate on this interpretation. First, further specification is needed to clarify the number of agents that are necessary and sufficient to make a claim about the realization of socialism true or false. Second, one also needs to specify whether quantity has priority over quality (or vice versa) when it comes to the agents. Consider the following two claims:

> (1) There is a collection of individual agents, namely the world's billionaires and political elites, that can, if working together, bring about socialism in the actual world.
>
> (2) There is a collection of individual agents, namely networks of ordinary socialist activists around the globe, that can, if working together, bring about socialism in the actual world.

It seems that the quality of the agents matters a great deal for whether socialism can or cannot be accomplished. The global elites seem much better placed to bring about

socialism than do ordinary socialist activists, given the level of money, time, political power and ideological influence required for the success of major political change. But quantity also matters, of course. It is one thing if there are just thousands of socialist activists around and quite another if it is half of the global population engaged in socialist activism.

There are, then, two parts to any claim about feasibility. First, there is the existential agent part – the issue of delimiting the existence, number and qualitative aspects of the agents who can intentionally bring about an outcome. Second, there is the ability part – the issue of providing an analysis of agents' abilities (also known as "agentive" or "agential" abilities). That is, an analysis of what it is for an agent to be able to do something intentionally. The second part is where political philosophers who have engaged in the nature of feasibility debate have focused most of their attention.[5]

It is important to emphasize that while feasibility claims entail agentive ability claims, they are not logically equivalent. For example, in English, statements about

[5] Indeed, much of the work on the nature of feasibility in political philosophy could easily be categorized as part of the literature on the metaphysics of abilities, although political philosophers have tended to prioritize the arguments and objections put forward by other political philosophers, instead of those advanced by metaphysicians. See e.g. Estlund (2016; 2020: 243-8); Gilabert and Lawford-Smith (2012: 810-8); Lawford-Smith (2012: 463-5; 2013); Southwood and Wiens (2016); Southwood (2016: 11-7); Stemplowska (2016; 2021); Wiens (2015; 2016). In turn, metaphysicians have largely ignored the claims about the nature of abilities coming from the literature on feasibility. For example, John Maier's (2022) entry on "Abilities" in *The Stanford Encyclopedia of Philosophy*, does not mention any of the work done by political philosophers in the last ten years.

agentive abilities can be paraphrased as feasibility statements using the expression "it is feasible for."[6] "Mary can go skydiving" can be paraphrased as "It is feasible for Mary to go skydiving." As Estlund (2020: 243) puts it, "[t]here is normally no error in speaking of what is feasible for an agent as simply a syntactic variant of discussing what the agent can, or is able to do."[7] Similarly, Gilabert and Lawford-Smith (2012: 812) write that "[c]laims about feasibility involve a four-place predicate concerning what a

[6] This seems to be what Zofia Stemplowska has in mind when she writes that "ability analyses of individual action are feasibility analyses too" (Stemplowska 2021: 2386). She also holds that "'Action Φ is feasible' means 'Action Φ is feasible for agent X'" (Stemplowska 2016: 290; 2021: 2386). But this is compatible with the idea that a state of affairs is feasible if there is at least one agent that can bring it about.

[7] Note, however, that this is not how Estlund chooses to use the term "feasibility" in his book *Utopophobia*. He explicitly refrains from providing a conceptual or philosophical analysis of "feasibility" and uses the term merely as an arbitrary label to denote the ability of non-agential groups (Estlund 2020: 243-4). His goal is to identify a notion that is analogous to the notion of individual ability in the principle of "ought implies can" (which applies to individual agents), but that, unlike individual ability, is suitable for non-agential groups. That is, the aim is to come up with a metanormative principle capable of assessing theories of justice in a manner similar to the way in which "ought implies can" serves as a standard for assessing moral theories. He finds that plural ability can play the role of individual ability. But, as result, claims about feasibility, which in his stipulative definition are claims about plural abilities, cannot be paraphrased as claims about ability, which he defines as claims about individual agents.

given agent can realistically do to accomplish something in a certain context" (i.e. "[i]t is feasible for *X* to φ to bring about *O* in *Z*").

However, claims involving the expression "it is feasible for" are different from the expression "Y is feasible." As Lawford-Smith (2013: 247) puts it, "[a]n outcome is feasible *iff* there exists an agent who can bring it about."[8] "Socialism is feasible" cannot be paraphrased by simply saying "X can bring about socialism" or "It is feasible for X to bring about socialism." For talking about X seems to come out of the blue. One needs to add the expression "there is" and an agent to the sentence. For example: "There is a group of agents X that can bring about socialism." Or "There is a group of agents X for

[8] She then elaborates on this as follows: "An outcome is feasible *iff* there exists an agent with an action in her (its) option set within the relevant temporal period that has a positive probability of bringing it about" (Lawford-Smith 2013: 250). For her, an agent *S* has the (agentive) ability to *F* if and only if *S* is likely to *F* if *S* tries (or chooses, intends, etc.) to *F* (Lawford-Smith 2013: 253; see also Gilabert and Lawford-Smith 2012). Gilabert and Lawford-Smith (2012) and Lawford-Smith (2013) distinguish between "binary" feasibility – *whether* a state of affairs is feasible – and "scalar" feasibility – *how* feasible a state of affairs is (see Wiens 2015 for some early criticism of this distinction and Silva 2019 for a reply). And Lawford-Smith (2013) argues that scalar feasibility is what matters in practical deliberation and in political philosophy. Note, however, that depending on one's analysis of abilities, the distinction between binary and scalar feasibility is redundant. For example, Romy Jaster argues that an agent has the (agentive) ability to ϕ "if and only if S ϕ's in a sufficiently high proportion of the relevant possible situations in which she intends to ϕ" (Jaster 2020, 95) and that abilities come in degrees, which can be expressed as the value of a fraction consisting of the set of relevant possible intention situations in which the agent ϕ's divided by the set of all possible relevant intention situations (Jaster 2020, 10, 96). So if Jaster is right, all feasibility is "scalar."

whom it is feasible to bring about socialism." And this means that feasibility claims in English are not logically equivalent to ability claims. They merely entail ability claims.

This is a purely semantic point, of course. But it is important for clarifying what we, English speakers, are talking about when we talk about feasibility. It gives us some direction regarding what the real substantive philosophical dispute about the nature of feasibility might be. It suggests the following question: can feasibility be analyzed in terms of the agents that exist and their abilities? Is it reducible to the agents that exist and their abilities?

The answer seems to be, yes. There is no need to postulate a property of feasibility that exists in its own right and that is primitive (i.e. unanalyzable), that should be simply taken at face value. Positing a property of feasibility in an actualist framework provides no extra explanatory power. It is unlike adding the property of possibility, entities such as states of affairs and propositions and the simple notion of truth, all of which make an important contribution to the actualist framework. If one is committed to the principle of ontological parsimony, one should simply eliminate the property of feasibility from one's ontology.

What if one eliminated, instead, agency and abilities from one's ontology? Then a claim such as "Socialism is feasible" might be about whether the state of affairs *socialism* bears the property of feasibility or not. Something would be needed to fill the vacuum left by the disappearance of agency and abilities and a property of feasibility would be able to do just that.

However, it is hard to see how an actualist (à la Plantinga) could reject the existence of agency and abilities to begin with. For example, those who deny the reality of agency and abilities on the grounds that only the atomic and subatomic particles postulated by contemporary physics exist (e.g. Ladyman and Ross 2007) would also deny the reality of a property of feasibility. Similarly, a possible worlds nominalist such as David Lewis (1973; 1986) would deny that agency and abilities exist in their own right, but would equally deny that any properties exist at all. These philosophers might concede that feasibility talk is helpful (just as one might concede that talk of properties is helpful) while still rejecting that such talk picks out a real and distinctive entity.

6. Why feasibility, understood as deliberation-worthiness, is not a property that exists in its own right

Perhaps one could argue for the existence of a property of feasibility by building upon Southwood's (2022) argument that feasibility is deliberation-worthiness. He provides not only a methodological approach to how we should theorize about feasibility, but also an account of what feasibility is. Southwood contends that (a) given that the role of our talk and thought about feasibility is to determine the deliberation-worthiness of actions and that (b) given that existing accounts of what feasibility is fail to limit the domain of what is feasible to the domain of the deliberation-worthy then (c) we have good reason to adopt an account of feasibility that identifies feasibility with deliberation-worthiness. This is the only watertight way, he claims, of guaranteeing that ascriptions of feasibility to actions will be neither too inclusive nor too exclusive and

pick out exclusively deliberation-worthy actions: after all, feasibility just *is* deliberation-worthiness, on this view.

Let us suppose that this account, which Southwood (2022, 150) calls the Fitting Deliberation Account of feasibility (FDA), is right and ignore my reservations about its methodological foundations. What does it imply for the ontological status of feasibility? Does it show that there is a property of feasibility that exists in its own right?

I do not think so. The FDA seems to be committed to only two major theses in metaphysics: first, feasibility is a functional property, not a modal property; second, feasibility is a property of action types, not states of affairs. This means that the FDA is fairly neutral with regard to many metaphysical controversies. It is perfectly compatible with nominalism about properties, which holds that properties do not exist in their own right and can be reduced, for example, to sets or classes of objects (e.g. Lewis 1986). In a nominalist framework, the FDA could be understood as identifying the class of action types that are deliberation-worthy. But it is also compatible with a realist position about properties, such as the version of actualism mentioned earlier. The FDA could be understood as identifying a *sui generis* property that some action types have, namely an irreducible property of deliberation-worthiness. This neutrality is a clear advantage of Southwood's account. Its fortunes are not tied up with controversial disputes in metaphysics. But this is of no help for settling the issue of whether there is a *sui generis* property of feasibility or not. One would need an additional argument, independent from the FDA, to defend that the latter implies a commitment to a *sui generis* property of feasibility.

In an actualist framework, the FDA can be clearly accommodated as an account of the existence of certain agents and their abilities to make a certain state of affairs obtain. Abilities here are understood in a strong agentive sense, as abilities to perform actions intentionally (cf. Jaster 2020, 13). Southwood (2022: 138; emphases added) comes close to suggesting this when he says that the claim that

> (1) "it is feasible (or infeasible) *for us* to bring about (or bring about and maintain) absolute equality of opportunity between men and women"

can be translated as

> (2) "it is feasible (or infeasible) *that* absolute equality of opportunity exists between men and women."

That is, he thinks that a claim about what is feasible for someone to do can be translated into a claim about the obtaining of a state of affairs (i.e. about what is feasible *simpliciter*) and vice versa. But he falls short of defending the claim that (2) can be translated into the claim:

> (3) There are (there are not) one or more agents that can bring about (or bring about and maintain) the state of affairs of absolute equality of opportunity between men and women.

I support the translation of (2) into (3) because, as already mentioned, figuring out what is feasible is just a matter of figuring out which agents are out there and what things they can do. Southwoodian feasibility is reducible to agentive abilities. There is no distinction, for example, between the principle of "ought implies can" and "ought

implies feasible," contrary to what Southwood (2016; 2022: 126) maintains. "I ought to do *X* only if it is feasible for me to do *X*" is the same as "I ought to do *X* only if I can do *X*" (again, I mean the agentive sense of "can").

When it comes to the FDA, I would say that if it is true, then the claim that "*X* is feasible" is just the claim that "There is at least one agent (or collection of agents) *F* who can *X* and for whom *X*-ing is deliberation-worthy." But there is no need to postulate a distinctive property of deliberation-worthiness here. One can explain deliberation-worthiness in terms of an obtaining state of affairs. For example, one could say that "*X*-ing is deliberation-worthy for *F*" amounts to the claim that the conjunction of states of affairs *F's not being pathologically afraid of X-ing*, *F's desiring, wanting, intending to X*, *F's having the opportunity and resources to X*, etc. obtain. There is no need to posit that, in addition to the action type/event/state of affairs *X* and the agent *F*, there is also a property of deliberation-worthiness. For example, for a person who has arachnophobia, the action of touching a spider is not worthy of deliberation. Here we don't need to postulate that there is a property of feasibility, which is absent when we think of the actions available to arachnophobic people but is present for ordinary people. We can just say that there are certain states of affairs that obtain or not. We can say that the state of affairs *being pathologically afraid of spiders* obtains or does not obtain.

Moreover, depending on the kind of action that is in question, the FDA becomes only trivially true. Any analysis of agentive abilities that is about temporally extended action (e.g. planning a holiday), as opposed to immediate actions (e.g. intentionally grabbing a falling mug before it hits the floor), entails that you can engage in the action only if you can deliberate about it. I cannot intentionally go on holiday to Paris if I never

deliberate about going on holiday to Paris. A modicum of deliberation is required of me so that I end up buying a plane ticket, packing my suitcase, calling a taxi and so on. Therefore, by definition, if you can engage in a temporally extended action, you can deliberate about it, and if you can deliberate about it, it is deliberation-worthy for you. So the claim that "*X* is feasible" amounts to the following claim is redundant:

> (A) There is at least one agent (or collection of agents) *F* who can *X* and for whom *X*-ing is deliberation-worthy.

(A) can be simply replaced, in the case of temporally extended actions, by:

> (B) There is at least one agent (or collection of agents) *F* who can *X*.

Again, this shows that even if the FDA is true, it does not seem to entail that there is a *sui generis* property of feasibility. There are only agents and their agentive abilities.

7. Conclusion

In sum, I have argued that there is no property of feasibility, given that feasibility is reducible to the agents that exist and the abilities that they have. That is the case even if it makes sense to identify feasibility with deliberation-worthiness.

Where does this leave the debate about the nature of feasibility? I believe that those interested in it should focus most of their attention on questions about the nature of agency and abilities. However, there is still interesting work to be done on feasibility itself. For example, there is the question of the number and character of agents required

to make an ascription of feasibility to a state of affairs true or false. For instance, there is something odd about saying that my receiving a billion dollars is feasible because there is at least one multi-billionaire in the actual world who can give me that kind of money. Perhaps an ascription of feasibility requires a minimal threshold concerning the number and quality of agents who can give me a billion dollars. Say, maybe it is feasible if and only if there is at least one multi-billionaire in this world *and* there is a number of coordinated groups of ordinary individuals working hard to fundraise the billion dollars on my behalf. This kind of question is not settled by merely looking at the nature of agency and abilities, so there is still room for the development of an analysis of feasibility.

References

Alvarez, Maria and John Hyman. 1998. "Agents and Their Actions." *Philosophy* 73 (284): 219-45.

Brennan, Geoffrey and Geoffrey Sayre-McCord. 2016. "Do Normative Facts Matter... To What Is Feasible?" *Social Philosophy and Policy* 33 (1-2): 434-56.

Chalmers, David. 2011. "Verbal Disputes." *Philosophical Review* 120 (4): 515-66.

Collins, Stephanie. 2019. *Group Duties: Their Existence and Their Implications for Individuals*. Oxford: Oxford University Press.

Collins, Stephanie. Forthcoming. "Abilities and Obligations: Lessons from Non-agentive Groups." *Erkenntnis*. https://doi.org/10.1007/s10670-021-00507-5

Davidson, Donald. 2001 [1967]. "The Logical Form of Action Sentences." In *Essays on Actions and Events*, ed. Donald Davidson, 105-22. 2nd edition. Oxford: Oxford University Press.

Erman, Eva and Niklas Möller. 2020. “A World of Possibilities: The Place of Feasibility in Political Theory.” *Res Publica* 26 (1): 1-23.

Estlund, David. 2016. “Reply to Wiens.” *European Journal of Political Theory* 15 (3): 353-62.

Estlund, David. 2020. *Utopophobia: On the Limits (If Any) of Political Philosophy*. Princeton: Princeton University Press.

Falguera, José, Concha Martínez-Vidal and Gideon Rosen. 2022. “Abstract Objects.” In *The Stanford Encyclopedia of Philosophy* (Summer 2022), ed. Edward Zalta. https://plato.stanford.edu/archives/sum2022/entries/abstract-objects/

Fine, Kit. 1976. Review: *The Nature of Necessity*, by Alvin Plantinga. *Philosophical Review* 85 (4): 562-6.

Fine, Kit. 1982. “First-Order Modal Theories III: Facts.” *Synthese* 53 (1): 43-122.

Fine, Kit. 2010. “Towards a Theory of Part.” *Journal of Philosophy* 107 (11): 559-89.

Gheaus, Anca. 2013. “The Feasibility Constraint on the Concept of Justice.” *Philosophical Quarterly* 63 (252): 445-64.

Gilabert, Pablo and Holly Lawford-Smith. 2012. “Political Feasibility: A Conceptual Exploration.” *Political Studies* 60 (4): 809-25.

Guillery, Daniel. 2021. “The Concept of Feasibility: A Multivocal Account.” *Res Publica* 27 (3): 491-507.

Hamlin, Alan. 2017. “Feasibility Four Ways.” *Social Philosophy and Policy* 34 (1): 209-31.

Jaster, Romy. 2020. *Agents’ Abilities*. Berlin: De Gruyter.

Korman, Daniel. 2020. “Ordinary Objects.” In *The Stanford Encyclopedia of Philosophy* (Fall 2020), ed. Edward Zalta. https://plato.stanford.edu/archives/fall2020/entries/ordinary-objects/

Ladyman, James and Don Ross (with David Spurrett and John Collier). 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Lawford-Smith, Holly. 2012. "The Feasibility of Collectives' Actions." *Australasian Journal of Philosophy* 90 (3): 453-67.

Lawford-Smith, Holly. 2013. "Understanding Political Feasibility." *Journal of Political Philosophy* 21 (3): 243-59.

Lawford-Smith, Holly. 2022. *Gender-Critical Feminism*. Oxford: Oxford University Press.

Lewis. David. 1973. *Counterfactuals*. Cambridge, MA. Harvard University Press.

Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Basil Blackwell.

Ludwig, Kirk. 2016. *From Individual to Plural Agency*. Oxford: Oxford University Press.

MacBride, Fraser. 2020. "Relations." In *The Stanford Encyclopedia of Philosophy* (Winter 2020), ed. Edward Zalta. https://plato.stanford.edu/archives/win2020/entries/relations/

Maier, John. 2022. "Abilities." In *The Stanford Encyclopedia of Philosophy* (Fall 2022), eds. Edward Zalta and Uri Nodelman. https://plato.stanford.edu/archives/fall2022/entries/abilities/

McLaughlin, Brian. 2003. "The Place of Colour in Nature." In *Colour Perception: Mind and the Physical World*, eds. Rainer Mausfeld and Dieter Heyer, 475-502. Oxford: Oxford University Press.

Orilia, Francesco and Michele Paoletti. 2022. "Properties." In *The Stanford Encyclopedia of Philosophy* (Spring 2022), ed. Edward Zalta. https://plato.stanford.edu/archives/spr2022/entries/properties/

Piñeros Glasscock, Juan and Sergio Tenenbaum. 2023. "Action." In *The Stanford Encyclopedia of Philosophy* (Spring 2023), eds. Edward Zalta and Uri Nodelman. https://plato.stanford.edu/archives/spr2023/entries/action/

Plantinga, Alvin. 1974. *The Nature of Necessity*. Oxford: Oxford University Press.

Plantinga, Alvin. 1976. "Actualism and Possible Worlds." *Theoria* 42 (1-3): 139-60.

Räikkä, Juha. 1998. "The Feasibility Condition in Political Theory." *Journal of Political Philosophy* 6 (1): 27-40.

Silva, Renan. 2019. "The Two-Dimensional Analysis of Feasibility: A Restatement." *Moral Philosophy and Politics* 6 (2): 357-78.

Southwood, Nicholas. 2016. "Does 'Ought' Imply 'Feasible?'" *Philosophy & Public Affairs* (44) 1: 7-45.

Southwood, Nicholas. 2018. "The Feasibility Issue." *Philosophy Compass* 13 (8): 1-13.

Southwood, Nicholas. 2022. "Feasibility as Deliberation-Worthiness." *Philosophy & Public Affairs* 50 (1): 121-62.

Southwood, Nicholas and David Wiens. 2016. "'Actual' Does Not Imply 'Feasible.'" *Philosophical Studies* 173 (11): 3037-60.

Stemplowska, Zofia. 2016. "Feasibility: Individual and Collective." *Social Philosophy and Policy* 33 (1-2): 273-91.

Stemplowska, Zofia. 2021. "The Incentives Account of Feasibility." *Philosophical Studies* 178 (7): 2385-401.

Textor, Mark. 2021. "States of Affairs." In *The Stanford Encyclopedia of Philosophy* (Summer 2021), ed. Edward Zalta. https://plato.stanford.edu/archives/sum2021/entries/states-of-affairs/

Wiens, David. 2015. "Political Ideals and the Feasibility Frontier." *Economics & Philosophy* 31 (3): 447-77.

Wiens, David. 2016. "Motivational Limitations on the Demands of Justice." *European Journal of Political Theory* 15 (3): 333-52.

Essay 2: Organizational Agency and the Argument from Empirical Social Science

Abstract

Are firms, political parties, states and other organizations, agents in their own right? That is, are they agents over and above the individuals that constitute them? While much has been written about this, little attention has been paid to the *argument from empirical social science*. This holds that one should believe in the reality of organizational agency if our most empirically successful social scientific research shows that organizational agency is indispensable for its success. I maintain that, as it stands, this argument fails, but that it could be potentially vindicated in future if it is taken in a new direction.

1. Introduction

Are organizations (things like states, firms, political parties, universities, trade unions, churches and so on) agents in the sense that human beings, sophisticated non-human animals such as dogs and monkeys and (more controversially) primitive single-cell organisms such as paramecia are? That is, do they *do* things – do they *themselves* engage in goal-directed behavior in any meaningful sense – or is talk of agency here merely a manner of speaking?

This question is often posed not only because many find it intrinsically interesting, but because philosophers usually think that the way one answers it has

implications for debates in moral, legal and political philosophy as well as for the way in which social science is practiced. For example, if states are considered to be intentional agents, then the idea of attributing moral responsibility directly to them, as opposed to individual politicians and civil servants, becomes plausible. Meanwhile, if it is conceptually incoherent to deny the agency of organizations, then social scientists might be ill-advised to do so, given that it might affect the explanatory or predictive power of their theories and models.

The dominant view in philosophy is that both coordinated social groups in general (which also include simple groups such as a couple walking together or a group of friends playing in a band), and organizations in particular, are agents in some sense.[1] However, most forms of realism about group agency, including naturalistic varieties,

[1] There are exceptions, of course, such as Ludwig (2016; 2017) and Moen (2023). For surveys of this debate, see e.g. Roth (2017) and Schweikard and Schmid (2021). For a representative sample of philosophical discussions of the existence and nature of organizations and organizational agency, see e.g. Tollefsen (2002); Isaacs (2011, ch. 1); List and Pettit (2011); Kaidesoja (2013, 316-9); Tuomela (2013, 231-40); Huebner (2014); Shapiro (2014); Epstein (2015); Ludwig (2017); Herzog (2018, ch. 4); Kincaid (2019, 187-91); Miller (2019); Little (2020); Strohmaier (2020a); Bratman (2022); Schweikard (2022); Collins (2023, chs. 1-3). I have excluded from this list works that focus on institutions (systems of rules) more broadly, instead of organizations specifically. Also, note that some works on group agency do *not* tackle organizational agency, often because of the very nature of the theory in question, which can only cope with simple and small groups (e.g. Kutz 2000; Gilbert 1989; Bratman 2014; for this point, see Shapiro 2014). For a social scientific survey of the history and nature of organizations, see e.g. Haveman (2022). For a theoretical account of organizational agency in international relations, see e.g. Gehring and Urbanski (2023).

such as those proposed by Tollefsen (2002) and List and Pettit (2011), have been quite abstract and supported by largely *a priori* arguments. While there is nothing wrong with that, there is a gap in the literature for more empirically grounded proposals.[2] Recently, Kincaid (2019) and List (2021) have made empirical cases for the reality of organizational agency, but they have done so only tentatively. This paper examines more closely the prospects for a largely empirical defense of realism about organizational agency – what one might call *the argument from empirical social science*. This draws upon a prominent view in the philosophy of science called "scientific realism," which states that we should accept as true (or approximately true) our best scientific theories and models and thus accept as real the entities and phenomena (both observable and unobservable) that such theories and models postulate.[3] So if our best social science postulates the existence of organizational agency, then we should accept that the latter is real.

[2] To be sure, there have been plenty of empirical studies on questions such as what happens in the brain when humans act together, how human cooperation evolved throughout the millennia or how ordinary people conceptualize joint action (see e.g. Tollefsen 2018, 396-8; Gomez-Lavin and Rachar 2019). But I am not talking here about a lack of empirical research on group agency, however defined. Rather, I am talking about the lack of empirically informed *arguments* for the existence of group agency and, specifically, organizational agency as a property that is irreducible to the properties of individual human beings.

[3] Kincaid (2019) does not frame his discussion of organizational agency in terms of scientific realism, but given that his argument is empirical and given his defense of scientific realism elsewhere (Kincaid 2000; 2018), it can be helpfully interpreted along scientific realist lines. List's (2021) scientific realist case for organizational agency is a special case of his more general scientific realist argument for the

I will argue that given the current state of the empirical scholarship, there is little reason to accept the argument from empirical social science. In the next section, I provide definitions for "organization," "agent," "scientific realism" and "empirical social science." Then, in section 3, I describe the argument from empirical social science in more detail. Section 4 attacks the argument on epistemic grounds while section 5 challenges it on a semantic basis. Section 6 maintains that the argument needs to follow a different strategy if it hopes to succeed in the future. Finally, section 7 concludes.

2. Definitions

Let me begin with some definitions. By an *organization*, I mean, following Collins (2023, 9) in most respects, a group of individuals "that involves a large number of people who realize a structure that coordinates divided labour via rules and hierarchical command relations, guided by a collective decision-making procedure." The only change I make to her definition is that where she uses the term "a collective agent," I use the term "a group of individuals." The reason for that is, of course, that my goal,

reality of intentional agency (List 2019, 56-8, 74-7). Note, though, that in earlier work with Philip Pettit he defended *interpretivism* about intentional agency (List and Pettit 2011; see also Strohmaier 2020b; Collins 2023, 21n7), which holds that intentional agency is the property of any system that is interpretable as an agent, and so intentional agency is "something that is in the eye of the beholder" (List 2019, 56).

unlike Collins', is to examine whether organizations can be agents or not, so it would beg the question if I included in the definition of organizations the claim that they are.

By an *agent*, I mean, following Barandiaran, Di Paolo and Rohde (2009, 369-73), a system that has at least the following three properties. First, it is distinct from its environment; second, it is capable of modulating some of its interactions with its environment – that is, it is not a passive object that is just kicked around by external forces; and, third, its interactions with its environment are governed by certain norms of what counts as a successful or failed interaction. The reason why I adopt this definition is because it is a fairly permissive notion of agency – even some single-celled organisms are considered agents according to it.[4] This minimal definition helps steelman the claim that the argument from empirical social science fails. After all, it is harder to deny that our most successful social science is committed to the view that organizations are minimal agents than it is to deny that organizations are fully fledged intentional agents with mental states. Moreover, more sophisticated forms of agency presuppose the properties of minimal agency, so once it is established that a system lacks them, questions about its having representational or intentional agency become moot.

When it comes to *scientific realism*, I follow Chakravartty's characterization of it (2017, introduction, §1.2):

[4] Burge (2009, 256-78; 2010, 327-41) provides an illuminating discussion of minimal agency, but he refrains from providing a definition of it. For a survey of the debate on minimal agency, see van Hateren (2022).

> Scientific realism is a positive epistemic attitude toward the content of our best theories and models, recommending belief in both observable and unobservable aspects of the world described by the sciences. […]
>
> Metaphysically, realism is committed to the mind-independent existence of the world investigated by the sciences. […]
>
> Semantically, realism is committed to a literal interpretation of scientific claims about the world. In common parlance, realists take theoretical statements at "face value". According to realism, claims about scientific objects, events, processes, properties, and relations […] whether they be observable or unobservable, should be construed literally as having truth values, whether true or false. […]
>
> Epistemologically, realism is committed to the idea that theoretical claims (interpreted literally as describing a mind-independent reality) constitute knowledge of the world.

A commitment to scientific realism does not require belief in all aspects of our best scientific theories. One need only be committed to those parts of the theories that merit belief (Chakravartty 2017, §1.3, §2.3; cf. Kincaid 2000). Described in this way, scientific realism seems quite intuitive and more plausible than alternative ways of thinking about science. For example, it is more compelling than constructive

empiricism (the view that we should only accept the observable entities/phenomena postulated by our best science), given that constructive empiricism fails to specify what in the social sciences could count as observable (Kincaid 2018, 373). To be sure, some philosophers have argued that even if scientific realism is true, the social sciences lack the sort of empirical success found in the natural sciences and so it is a bad idea to try to settle metaphysical disputes about the nature of social reality by appeal to empirical research (e.g. Hawley 2018, 189-92). However, there *have* been fairly successful instances of empirical social science, as I mention below. In any case, for the purposes of this article, I ask readers to grant me the assumptions that scientific realism is true and that it makes sense in the context of the social sciences. My concern here is with whether, granted those assumptions, the social sciences vindicate the reality of organizational agency.

Finally, let me clarify what I mean by "empirical social science." By that I mean the investigation of social phenomena via the collection and analysis of empirical data, which includes data from government records, consumer data, surveys, interviews and lab and field experiments. This stands in contrast with "theoretical social science," which involves mostly *a priori* reasoning. Social theory, non-applied decision, game and social choice theory, philosophy of social science, philosophy of collective agency and much of analytic social metaphysics/ontology all fall within the category of theoretical social science. Although most of the social scientific research on the existence of organizational agency is largely theoretical, not all theoretical research on it is social scientific. For example, some arguments for the existence of organizational agency do not claim to be advancing the social sciences, theoretical or empirical. They

are, instead, intended to be contributions purely to debates in, say, ethics or metaphysics (e.g. Rovane 1998).

3. The argument from empirical social science

According to the *argument from empirical social science*, we should believe that organizational agency exists because our most empirically successful social scientific theories and models require the postulation of organizational agency. More specifically, we should accept that at least some organizations are agents because of the empirical success of rational choice explanations and predictions in economics and political science, which posit that firms, states and political parties (among other organizations) are agents. Such empirically successful explanations/predictions include explanations/predictions of the behavior of firms in microeconomics, explanations/predictions of the behavior of political parties that draw upon expected utility theory and explanations/predictions of the behavior of states in the international arena that draw upon game theory (List 2021, 1215-6; cf. Kincaid 2019, 189-90). As List (2021, 1215-6) puts it,

> [o]ur best social-scientific theories represent some collective entities as goal-directed agents and explain their behaviour by using the same concepts and categories that we use to explain individual behaviour. For example, the theory of the firm in economics and "realist" theories of international relations apply standard rational-actor models to firms and states. In fact, a profit-

> maximizing firm may be a more fitting case of a *homo economicus*, a self-interested utility-maximizing rational agent, than any individual human being is. And strategic interactions between states, such as between the USA and the Soviet Union during the Cold War, are often modelled as games with strategically rational players. Similar points apply to the way political scientists think about parties and other organizations in politics.

Although rational choice theory is usually understood as "a theoretically more sophisticated, mathematical version of our folk psychological theory of intentional action common in philosophical action theory" (Herfeld 2022, 2) – which holds that an action is behavior that is caused by a combination of beliefs, desires and intentions – rational choice theory is neither a single theory nor is it necessarily committed to the notions of *intentional* action or agency (Herfeld 2022). So the empirical case for organizational agency can be broadened and include not only explanations that presuppose the presence of intentional agency (e.g. List 2021) but also those that only appeal to non-intentional agency (e.g. Kincaid 2019). For instance, Kincaid (2019) argues that microeconomics provides a good example of a non-intentional agential (or "behaviorist") explanation that is committed to organizational agency. He writes:

> There is a theoretically well-developed and empirically well-supported body of work in microeconomics on the theory of the firm. Firms are treated as unitary agents in a way parallel to the way consumers are treated in revealed preference theory: They

> have in effect a utility function where the argument is usually profit maximization but can also include other things. They face constraints similar to the consumer budget constraint but in the form of production possibilities and the costs of inputs and the externally determined prices of outputs. From observed choices of inputs and of prices for outputs it is possible to describe firm behavior in revealed preference fashion: The internal workings of the firm can be ignored and yet still the observed behavior explained as a function of external constraints. (Kincaid 2019, 189-90)

While one should broaden the argument from empirical social science beyond intentional agency to include agency more generally, it is also important to narrow it down in one respect. It should focus exclusively on *empirical* explanations – that is, explanations that rely on empirical data. An empirical argument for organizational agency should not be conflated with more general indispensability arguments for organizational agency, which include empirical *and* purely theoretical or philosophical considerations. For example, List (2021, 1216) provides an indispensability argument, as opposed to an empirical argument:

> **Premise 1**: The ascription of intentional agency to certain organized collectives is explanatorily indispensable if we wish to make sense of their behaviour.
>
> **Premise 2**: If the ascription of some property to an entity is indispensable for explaining that entity's behaviour, then we have

a provisional justification for assuming that the entity really has that property.

Conclusion: We have a provisional justification for assuming that the collectives in question really have the property of intentional agency.

As List (2021, 1216) himself observes, "[p]remise 1 is a partly empirical and partly methodological claim about the social sciences." However, I believe this kind of general indispensability argument can muddy the waters, if one is not careful. Just because a theoretical or philosophical argument for organizational agency succeeds does not mean that an empirical argument also succeeds. One should keep philosophical/theoretical and empirical indispensability arguments separate from each other both if one is committed to the view that philosophy is continuous with science and if one is not. For example, someone who is convinced by David Lewis' (1986) philosophical indispensability argument for the existence of a plurality of worlds would need to find a way to reconcile that with the fact that the findings of experimental physics are based on the assumption that there is only our world. A Lewisian would have to say that it is a mistake to try and settle the question of whether it is explanatorily indispensable *in general* that we postulate a plurality of worlds. He or she would have to maintain that establishing whether it is theoretically or philosophically indispensable to postulate the existence of a plurality of worlds is a different kind of project to that in which experimental physicists are engaged.

However, my guess is that, given their naturalistic inclinations, most theorists in the organizational agency debate would want indispensability arguments for organizational agency to be continuous with empirical research. So the claim that postulating group agency is explanatorily indispensable would be akin to that of theoretical physicists in the 1960s who theorized that there was such a thing as a Higgs boson, but whose existence was only empirically demonstrated in 2012 (see e.g. Salam et al. 2022). So just as it made sense to distinguish the empirical research on the Higgs boson from the theoretical research (even though the former depended on the latter), it is helpful to separate theoretical from empirical inquiry on organizational agency. The fact that most philosophers believe that organizations can be agents (List 2018, 295) tells us nothing about the empirical support for organizational agency and so one should distinguish between theoretical and empirical indispensability arguments even if one is fully committed to a scientific worldview.

Given all of the above, I believe that an empirical argument for the reality of organizational agency should look something like this:

1. Scientific realism is true. That is, we should accept as true (or approximately true) our best scientific theories and models and thus accept as real the entities and phenomena (both observable and unobservable) that such theories and models postulate.
2. Scientific realism is selective. That is, a commitment to scientific realism need not entail a belief in the truth of *all* parts of our best scientific theories and models.

3. One should believe in those parts of a theory or model that are empirically successful in the sense that they are mature, non-*ad hoc* and possess a high degree of explanatory and/or predictive power.[5]
4. Some parts of our best social scientific theories and models are empirically successful because they postulate that some organizations are agents.
5. Therefore, we should believe that organizations can be agents.

4. The epistemic challenge to the argument from empirical social science

Let us consider the examples suggested by List and Kincaid, which allegedly support the empirical case for organizational agency. I will argue that we lack reasons to believe in the empirical success of these examples and that believing in their empirical success does not entail believing in the reality of organizational agency.

First, there is the theory of the firm, which is the study of the nature and behavior of firms. Here it is important to distinguish between theoretical and applied research. In the theoretical domain, there has been little progress in the last 20 years or so (Hodgson 2019, 220). Economists have proposed radically different definitions of the word "firm," with the only apparent agreement being that it refers to "some site where production takes place" (Hodgson 2019, 222; see also Hodgson 2015a, ch. 8). Crucially, none of the major figures in this debate have defined the firm as anything that might resemble an agent (see Hodgson 2015a, ch. 8). So why do List and Kincaid suggest that the theory of the firm supports the empirical case for organizational agency? I believe it is because they are focusing on what economists *do*, instead of what they say. It is

[5] This draws on Chakravartty (2017, §1.3, §2.3).

common in economics to ascribe utility functions to firms and given that any entity that has a utility function behaves like an agent (i.e. like a system with goal-directed behavior), then it seems plausible to treat firms as agents. So the argument for organizational agency seems to be that because microeconomics has been relatively successful in predicting the behavior of firms given internal changes in the firms and shifts in the wider economy, it seems that we should be realists about the claims made by microeconomics, including the claim that firms are agents.

However, it is unclear why the ascription of utility functions to firms should leads us to be realists about the idea that firms are agents, instead of leading us to treat firms only *as if* they were agents. First, recall that scientific realism is not a commitment to believing in the truth of *all* parts of a scientific theory or model. One can be a realist about much of microeconomics without believing in the reality of firm agency. The equation of agency with the possession of a utility function is a theoretical claim (cf. Hodgson 2015a, 219). One can accept the usefulness of ascribing utility functions to firms without accepting that that ascription is evidence that firms are agents. Second, the empirical data in economics does not determine which theories of the firm should be accepted as true. Whether one should side with the majority of economists, who reject that firms are agents, or side with those theorists and philosophers who believe in firm agency cannot be settled by looking at the empirical data. The view that firms are paradigms of the *homo economicus* (List 2021, 1215-6) cannot be inferred from the findings of microeconomics. Rather, it is just one interpretation of the data among many, which, given the current state of empirical research, can only be settled, if at all, on theoretical grounds.

Now consider rational choice explanations of the behavior of political parties and states. Take states first. The most influential rational choice theory of international relations is "structural realism" (Waltz 1979), which attempts to predict and explain the behavior of states by treating them as utility-maximizing agents that act partly in response to the pressures of the international state system. (Note that "realism" in international relations has a different meaning to "realism" in philosophy.) It is dubious, to say the least, that structural realism has been empirically successful. As many scholars have noted, it is striking how miserably it failed to predict and explain the unexpected end of the Cold War. While structural realists insist that "the Cold War ended exactly as structural realism led one to expect" (Waltz 2000, 39), it is widely believed that such claims are *post hoc* rationalizations and that structural realism has failed in its aims (see e.g. Lebow 1994).

To be sure, structural realism is just one realist theory of international relations (Wohlforth 1994, 92). However, it is unclear that other realist theories and, indeed, non-realist rational choice theories of international relations (e.g. liberal approaches), have been empirically successful in a mature and non-*ad hoc* way. For example, one of the most important and enduring debates in international relations has been about whether economic interdependence leads to peace or war (Copeland 2014). Andrew Moravcsik (1997; 1998) provides one of the most influential rational choice liberal theories of international relations, according to which the behavior of states is to a large extent determined by the preferences of individuals and interest groups within those states. He used this theory to explain the remarkable phenomenon of European integration – that is, the fact that after the two bloodiest conflicts in human history, former European

rivals got together and formed a peaceful union, the European Union (EU) (Moravcsik 1998). Moravcsik (1998) argues that European integration was caused by economic interdependence. However, while most scholars agree with him that this is a good explanation for how European integration developed in the first forty years of the EU's history, it is highly controversial what theory best explains the EU's trajectory since 1992 (Moravcsik 2018, 1649). There is a clear parallel here with the Downsian approach to American politics (more on this shortly). Both Downs and Moravcsik developed general theories that seemed plausible at first but turned out to struggle once circumstances, not foreseen by the original theory, changed.

The debate about the effects of economic interdependence on international relations is further complicated by the fact that there is robust statistical evidence to support both the liberal claim that economic interdependence makes peace more likely *and* the realist claim that it makes *war* more likely (Copeland 2014, 24-5). Nowadays, it is widely acknowledged that there is a degree of truth to both claims: economic interdependence makes peace more likely in some contexts, but war more likely in other contexts (Copeland 2014, 24). However, the challenge remains of explaining why this is the case (Copeland 2014, 25). In short, there is a lack of mature and non-*ad hoc* theories to explain (and predict) the effects of economic interdependence on war and peace, so we cannot turn to this field of study – the oldest and most studied in the discipline of international relations – to support the case for believing in the reality of organizational agency.

Furthermore, even if we accepted that rational choice approaches to international relations *do* have great predictive or explanatory power, we would still face the issue

that such approaches could be plausibly interpreted as merely treating states *as if* they were agents. For example, in structural realism, the assumption that states are utility-maximizing agents is just one element of the theory and not necessarily the most important one (Goddard and Nexon 2005), so it is perfectly legitimate to take that assumption as a helpful idealization or fiction, while still accepting at face value – that is, accepting as accurate descriptions of the world – the other parts of the theory. Indeed, as Wendt (2004, 289-90) observed twenty years ago (and as it is still true today), most IR theorists do not believe that states are *literally* agents. So one would need a pretty good reason to successfully challenge IR scholars' own interpretation of their theories.

Now let us turn to rational choice theories of party behavior. The most famous of them is Downs (1957), who dominated the study of American politics until the beginning of this century. He argued, among other things, that political parties in a system like the American one appeal to the median voter because this is how office-seeking politicians survive in a competitive electoral environment (for a prominent revised version of the Downsian approach, see Aldrich 1995). Here political parties themselves can be modelled as utility-maximizing agents facing evolutionary pressures to act rationally, which otherwise face the prospect of extinction. Even if Downs' theory of party behavior had been empirically successful, there would be little reason to suppose that we should accept at face value the claim that parties are literally agents. After all, all that would be needed for successful prediction and explanation is that parties should be treated *as if* they have certain preferences, such as a preference for survival (Satz and Ferejohn 1994, 79; Herfeld and Marx 2023, 71-2).

However, the problem is that the theory has *not* been empirically successful, neither in terms of prediction nor explanation, as critics have long pointed out (Green and Shapiro 1994). Most obviously, there is the failure of the median voter theorem. While it might have reliably predicted electoral outcomes in the United States back in the 1950s, it has been unable to predict electoral outcomes in recent decades. In addition to struggling with prediction, the Downsian framework has been unable to explain two dramatic developments in American politics – the emergence of asymmetric polarization and the steady rise in income inequality (Hacker and Pierson 2014, 652-5). Asymmetric polarization, in the US context, refers to the fact that while both the Democratic and Republican parties have moved to opposite extremes of the ideological spectrum, Republicans have moved further to the right than Democrats have moved to the left and yet the former have not faded into electoral irrelevance (Hacker and Pierson 2014, 652-3; Pierson and Schickler 2020, 51). According to the Downsian view, this is baffling. Downsians explain political outcomes by focusing on voters' preferences and elections but these have been unable to fill the explanatory gap. As Hacker and Pierson (2014, 653) observe,

> The Downsian approach has a very hard time making sense of such a persistent and lopsided departure from the center in a competitive party system. After all, there is nothing in public-opinion data suggesting a similar move to the right among voters on central policy issues. To Downsians, the continuing right turn of the GOP poses something of an existential puzzle: Does

electoral competition really matter so little for the positioning of the parties?

Survey data has consistently shown that, contrary to popular perceptions, American voters are *not* extremely polarized in their ideological positioning and so the high degree of polarization among elected officials has to be explained by something other than voters' preferences (Fowler et al. 2023, 658-9).[6]

Similarly, Downsians struggle to explain why there has not been a backlash among average American voters with regard to their falling relative income over the past few decades (Hacker and Pierson 2014, 654-5). If the Downsian framework is right, voters will care that they are worse off and demand that politicians do something about it and politicians, in turn, will be responsive to these demands. However, this is not what has happened. As a result, even some prominent Downsian political scientists have started to move beyond the Downsian framework in order make sense of this phenomenon (e.g. Bartels 2008; see also Hacker and Pierson 2014, 654).

The failure of the Downsian approach has given impetus to an alternative framework, which some scholars have called the "Schattschneiderian" or "policy-focused" framework (Hacker and Pierson 2014, 644). This draws directly or indirectly on the work of E. E. Schattschneider (1935; 1942; 1960), who, unlike Downs, emphasized the importance of interest groups in shaping politics. One influential

[6] Campbell (2016) challenges this by arguing that American voters *are* polarized and *they*, as opposed to political elites, are the cause of polarization in American politics. However, as Levendusky (2018) argues, Campbell's argument is unconvincing and relies on an idiosyncratic conception of "polarization," which is not shared by most scholars of American politics.

application of the Schattschneiderian framework has been the so-called "UCLA model," which until recently had become "a new conventional wisdom, providing a lens for understanding American politics with great influence both in political science and in academically informed journalistic analyses" (McCarty and Schickler 2018, 176). According to this model, we should revise our definition of political parties in the United States and take a party to be constituted not only by the formal party itself, but also by the interest groups and activists surrounding it. Elected politicians and party officials are taken to be mere delegates enacting the interests of such powerful groups and presidential primaries are seen as rubber-stamping exercises that ratify what has already been agreed upon in private by the elites (Cohen et al. 2008; Bawn et al. 2012). Crucially, for our purposes, on this broader understanding of what parties are, an American political party is still modelled as a rational agent with preferences irreducible to those of the individuals and groups constituting it (McCarty and Schickler 2018, 177). However, the UCLA model has been an empirical failure in at least two significant respects: it predicted the victory of Hilary Clinton in the 2008 Democratic primaries and the defeat of Donald Trump in the 2016 presidential election (McCarty and Schickler 2018). This does not mean, of course, that the wider Schattschneiderian framework has been a failure. The jury is still out. Scholars of American politics, which is one of the most intensely studied topics of political science, are still reeling from the fact that these events took place as well as from the rise of extreme polarization and soaring income inequality in the US. So whatever plausible theories and models come out of this field of study in the foreseeable future will not be mature and non-*ad hoc* enough to support the case for realism about organizational agency.

Leaving aside microeconomics, realist approaches to international relations and theories of party politics, what other potential candidates for empirically successful social science are there? Some might point as examples democratic peace theory, which is the study of why democratic states do not go to war with each one another (see e.g. Bueno de Mesquita and Smith 2012, 166-70); research on the income inequality hypothesis in social epidemiology, which seems to successfully demonstrate a strong relationship between levels of income inequality and health (Saunders 2020, 146-52); election prediction via opinion polling (Northcott 2015; 2017); and the study of traffic flows (Jusup et al. 2022, 11-6). Although these seem to be, indeed, rare success stories in the social sciences in the sense of being empirically well-supported, mature and non-*ad hoc* theories and models, none of these examples appeal to organizational agency.

5. The semantic challenge to the argument from empirical social science

So far, I have been focusing on the epistemic dimension of scientific realism. Specifically, I have examined whether we should believe in the truth of the claims made by our best social science. However, problems for the argument from empirical social science also arise in the *semantic* dimension. Even if we accepted as true all the claims made by microeconomics, structural realism in international relations and Downs' theory of party behavior – that is, as claims about the world, as opposed to mere idealizations or useful fictions – a semantic analysis of these claims would not necessarily commit us to the interpretation that we are literally saying that firms, states and political parties are agents. Linguistic data shows that a sentence such as "Firms raise prices based on supply and demand" does not necessarily mean that the firms are

the agents of the event of raising prices. As Ludwig argues (2017, esp. chs. 4, 5, 13), action sentences (e.g. "X did Y") involving singular group referring terms (e.g. "General Electric," "Congress," "France") in the subject position can be interpreted following what he calls the "multiple agents analysis." According to this analysis, a sentence such as "General Electric laid off 10% of its workforce" should be interpreted as saying (roughly) that there was an event in which the directors of General Electric, acting on behalf of its shareholders, laid off 10% of its employees.

Even if one is not prepared to accept Ludwig's (2017) semantic analysis, one should at least acknowledge that collective action sentences are ambiguous. As Fleming (2017, 938-44) observes, sentences such as "France conducted airstrikes" have three possible meanings. First, one could interpret "France" as an agent that went on an airplane and itself carried out airstrikes. Second, one could interpret the sentence as saying that "France" is an agent in its own right but that it itself did not carry out the airstrikes – it only authorized others to do so. Or, third, one could read it as saying that the state of France is causally and/or morally responsible for the airstrikes but that it is not an agent, so it neither conducted the airstrikes itself nor did it authorize them – some French pilots conducted the airstrikes authorized by some French politicians. Like Ludwig, Fleming (2017) believes that English speakers do not attribute agency to states. Rather, English speakers take individual agents to act on behalf of the state. Regardless of the persuasiveness of this argument, it is a fact that collective action sentences have different readings. Therefore, even if our best microeconomic theories, structural realism in IR, etc. were true, we could be committed to the truth of the sentences

contained in those theories without being committed to the view that those sentences imply that organizational agency is real.

However, one might wonder, what about those social scientists who explicitly state that the organizations in their theories and models are literally agents? Surely, one could argue, Ludwig's and Fleming's semantic analyses would be putting words in the social scientist's mouth, distorting what the latter is saying. For instance, consider Douglass North's work on economic development. This is a clear example of empirical research that openly states that organizations are agents and so it might seem odd to apply a semantic analysis to it that contradicts that. North developed a highly influential theory of institutional change in order to argue that different societies show different levels of economic development because of the varied ways in which specific institutions and organizations have shaped one another in the various economies across both time and space (North 1990). He adopted the traditional definition of institutions as systems of rules, but he claimed that organizations are not a type of institution, but something else (North 1990, 3-5).[7] For North, organizations "are groups of individuals bound by some *common purpose* to achieve objectives" and "in the course of attempts to accomplish their objectives are a major *agent* of institutional change" (1990, 5;

[7] This puts him in disagreement with theorists such as Bratman (2022), Hodgson (2006; 2015b; 2017) and others. For Hodgson, for example, "[o]rganizations are special institutions that involve (a) criteria to establish their boundaries and to distinguish their members from nonmembers, (b) principles of sovereignty concerning who is in charge, and (c) chains of command delineating responsibilities within the organization" (Hodgson 2006, 8). Similarly, Bratman refers to organizations as "organized institutions" (Bratman 2022, xi), but he, unlike Hodgson, believes that organizations can be agents.

emphases added). The distinction between institutions and organizations, he says, is analogous to that between rules and players in a game:

> Conceptually, what must be clearly differentiated are the rules from the players. The purpose of the rules is to define the way the game is played. But the objective of the team within that set of rules is to win the game - by a combination of skills, strategy, and coordination […]. Modeling the strategies and the skills of the team as it develops is a separate process from modeling the creation, evolution, and consequences of the rules. (North 1990, 4-5)

Later on, he and two colleagues elaborated on the nature of organizations as follows (North, Wallis and Weingast 2009, 15-6):

> *organizations* consist of specific groups of individuals pursuing a mix of common and individual goals through partially coordinated behavior. Organizations coordinate their members' actions, so an organization's actions are more than the sum of the actions of the individuals. Because they pursue a common purpose in an organization and because organizations are typically composed of individuals who deal with each other repeatedly, members of most organizations develop shared beliefs about the behavior of other members and about the norms or rules of their organization. As a result, most organizations have their own internal institutional structure: the rules, norms, and

> shared beliefs that influence the way people behave within the organization (Greif, 2006).

So it is clear that North (1990) and North, Wallis and Weingast (2009) are committed to the idea of organizational agency. And given that these works (like much research in political economy) shy away from idealizations and useful fictions, we can be confident that the claims made by North and his colleagues should be taken at face value. Nonetheless, the meaning of the collective action sentences uttered by them, such as "organizations incrementally alter the institutional structure" (North 1990, 73), is still ambiguous here. To see why, we need to distinguish between the *speaker-meaning* and the *semantic-meaning* of a linguistic expression. While the speaker-meaning of North et al.'s collective action sentences is not ambiguous, the semantic-meaning is. Pinder (2021, 152) helpfully explains the distinction as follows:

> *Speaker-meaning* is what a *speaker* means when she utters some words. It is closely tied to a speaker's intentions: in ordinary conversational exchanges, what a speaker means when she utters some words is just what she intends to convey by her utterance. In contrast, *semantic-meaning* is what the uttered *words* mean. For concreteness, we can think of semantic-meaning as closely tied to the idea of a linguistic community: what a word means is governed in some way by the linguistic community to which the speaker belongs.

So by the sentence "organizations incrementally alter the institutional structure" (North 1990, 73), North intends to convey that certain kinds of complex group agents, such as firms, incrementally alter the institutional structure. By contrast, what this sentence actually means in the linguistic community of English-speaking social scientists is ambiguous, because the word "organizations" can be given both an agential and a non-agential reading. And, again, whether one settles on an agential or non-agential reading is a theoretical choice. So we could, in principle, accept that North's work is an empirical success and thereby be committed to the truth of all his claims, such as that "organizations incrementally alter the institutional structure," without accepting that the words "organizations," "firms" and so on refer to real group agents.

It should be noted that the semantic challenge to the argument from empirical social science only has force if one believes that a philosophical argument can be undermined by semantic considerations. Or if one believes that, other things being equal, it is legitimate to prefer one philosophical argument instead of another because of the former's non-philosophical virtues, such as its simplicity, elegance or compatibility with our best semantic theories. So if one believed that North's theory of institutional change is an empirical success, then one would have a strong case for believing that organizations can be agents. However, the case would not be as strong as it could otherwise be because there are semantic problems. One can draw an analogy here with someone going to a dealership to buy a car to use for daily activities, like going to work. It should be enough for the buyer if the car does its job well of moving the driver from place to place, but it counts in favor of a specific car if it can do more than that – if it looks good or has comfortable seats, say.

6. What would vindicate the argument from empirical social science?

The argument from empirical social science has failed so far for two reasons. First, there is a lack of empirically successful social scientific theories and models that refer to organizations as agents. Second, even the theories and models that *do* refer to organizations as agents can be interpreted as merely treating organizations *as if* they were agents. How could one tackle these challenges?

I believe the most promising strategy that someone interested in this line of inquiry can follow is to focus on theories and models in which organizational agency is the key *independent variable* explaining the occurrence of a phenomenon. That is, the focus should move away from theories and models that merely talk about organizations as agents to those in which empirical success is the direct result of postulating organizational agency as the main cause of a phenomenon. The theories and models considered above – microeconomic models of firm behavior, structural realism in IR, theories of American party politics and Douglass North's theory of institutional change – all make use of organizational agency talk, but none claims that it is the very presence of organizational agency that uniquely predicts or explains the phenomenon under consideration. It is one thing to say that political parties should be treated as agents in a model and quite a different thing to say that the agency that parties have predicts and explains their behavior. If party agency is the key independent variable of a theory or model and the latter is empirically successful, then we have a vindication of the empirical argument for organizational agency. Consider another example. We could develop a study investigating what predicts or explains the differing levels of military expenditure by states by looking at whether and how different forms of organizational

agency affect military spending. If studies of this kind demonstrate that organizational agency is a key independent variable and they turn out to be empirically successful in the long-run, then there would be a strong reason to accept the reality of organizational agency.

Fortunately, one would not have to start from scratch here. There is already a vast empirical literature on variables similar to the three properties described by Barandiaran, Di Paolo and Rohde (2009, 369-73) in their definition of agency, such as how the autonomy of organizations, organizational capacity and organizational norms explain certain phenomena. For example, there have been studies arguing that the level of central bank independence affects the unemployment level, credit conditions and the stock market during banking crises (Hansen 2022); studies arguing that state capacity (e.g. the ability of a state to provide public goods such as education and security) affects levels of democratization in postcommunist countries (Fortin 2012) and studies arguing that norms impact how civil wars end (Howard and Stark 2018). The challenge would be to demonstrate whether tying together these variables in any single study would have a greater predictive or explanatory power than those studies that consider those variables separately.

7. Conclusion

In summary, there are epistemic and semantic reasons to doubt that our best empirical social science lends support to the thesis that organizations can be agents. Either the putative cases of empirical support for organizational agency turn out, on closer inspection, to be empirical failures or they do not require a commitment to the reality

of organizational agency at all. If the argument from empirical social science is to succeed, one should turn to theories/models that treat organizational agency as a key independent variable – a central cause of an event, process or pattern. It is fruitless to focus on studies in which organizational agency is just one assumption among many, for here such an assumption can be interpreted as a mere idealization or fiction or just as a manner of speaking. And if one starts trying to refute claims that such organizational agency talk is more than just talk, one is back in the domain of largely *a priori* theoretical or philosophical research and no longer providing an argument from empirical social science.

References

Aldrich, John. 1995. *Why Parties? The Origin and Transformation of Political Parties in America*. Chicago: University of Chicago Press.

Barandiaran, Xabier, Ezequiel Di Paolo and Marieke Rohde. 2009. "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action." *Adaptive Behavior* 17 (5): 367-86.

Bartels, Larry. 2008. *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton: Princeton University Press.

Bawn, Kathleen, Martin Cohen, David Karol, Seth Masket, Hans Noel and John Zaller. 2012. "A Theory of Political Parties: Groups, Policy Demands and Nominations in American Politics." *Perspectives on Politics* 10 (3): 571-97.

Bratman, Michael. 2014. *Shared Agency: A Planning Theory of Acting Together*. New York: Oxford University Press.

Bratman, Michael. 2022. *Shared and Institutional Agency*. New York: Oxford University Press.

Bueno de Mesquita, Bruce and Alastair Smith. 2012. "Domestic Explanations of International Relations." *Annual Review of Political Science* 15: 161–81.

Burge, Tyler. 2009. "Primitive Agency and Natural Norms." *Philosophy and Phenomenological Research* 79 (2): 251-78.

Burge, Tyler. 2010. *Origins of Objectivity*. New York: Oxford University Press.

Campbell, James. 2016. *Polarized: Making Sense of a Divided America*. Princeton: Princeton University Press.

Chakravartty, Anjan 2017. "Scientific Realism." In *The Stanford Encyclopedia of Philosophy* (Summer 2017), ed. Edward Zalta.

https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/

Cohen, Marty, David Karol, Hans Noel and John Zaller. 2008. *The Party Decides Presidential Nominations Before and After Reform*. Chicago: University of Chicago Press.

Collins, Stephanie. 2023. *Organizations as Wrongdoers: From Ontology to Morality*. New York: Oxford University Press.

Copeland, Dale. 2014. *Economic Interdependence and War*. Princeton: Princeton University Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.

Epstein, Brian. 2015. *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. New York: Oxford University Press.

Fleming, Sean. 2017. "Artificial Persons and Attributed Actions: How to Interpret Action-Sentences about States." *European Journal of International Relations* 23 (4): 930–50.

Fortin, Jessica. 2012. "Is There a Necessary Condition for Democracy? The Role of State Capacity in Postcommunist Countries." *Comparative Political Studies* 45 (7): 903-30.

Fowler, Anthony, Seth Hill, Jeffrey Lewis, Chris Tausanovitch, Lynn Vavreck and Christopher Warshaw. "Moderates." *American Political Science Review* 117 (2): 643-60.

Gehring, Thomas and Kevin Urbanski. 2023. "Member-dominated International Organizations as Actors: A Bottom-up Theory of Corporate Agency." *International Theory* 15 (1): 129-53.

Gilbert, Margaret. 1989. *On Social Facts*. Princeton: Princeton University Press.

Goddard, Stacie and Daniel Nexon. 2005. "Paradigm Lost? Reassessing *Theory of International Politics*." *European Journal of International Relations* 11 (1): 9–61.

Gomez-Lavin, Javier and Matthew Rachar. 2019. "Normativity in Joint Action." *Mind & Language* 34 (1): 97-120.

Green, Donald and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science.* New Haven: Yale University Press.

Hacker, Jacob and Paul Pierson. 2014. "After the 'Master Theory': Downs, Schattschneider, and the Rebirth of Policy-Focused Analysis." *Perspectives on Politics* 12 (3): 643-62.

Hansen, Daniel. 2022. "The Economic Consequences of Banking Crises: The Role of Central Banks and Optimal Independence." *American Political Science Review* 116 (2): 453-69.

Haveman, Heather. 2022. *The Power of Organizations: A New Approach to Organizational Theory*. Princeton and Oxford: Princeton University Press.

Hawley, Katherine. 2018. "Social Science as a Guide to Social Metaphysics?" *Journal for General Philosophy of Science* 49 (2): 187-98.

Herfeld, Catherine. 2022. "Revisiting the Criticisms of Rational Choice Theories." *Philosophy Compass* 17 (1): 1-20.

Herfeld, Catherine and Johannes Marx. 2023. "Rational Choice Explanations in Political Science." In *The Oxford Handbook of Philosophy of Political Science*, eds. Harold Kincaid and Jeroen Van Bouwel, 54-85. Oxford: Oxford University Press.

Herzog, Lisa. 2018. *Reclaiming the System: Moral Responsibility, Divided Labour, and the Role of Organizations in Society*. New York: Oxford University Press.

Hodgson, Geoffrey. 2006. "What Are Institutions?" *Journal of Economic Issues* 40 (1): 1-25.

Hodgson, Geoffrey. 2015a. *Conceptualizing Capitalism: Institutions, Evolution, Future*. Chicago: University of Chicago Press.

Hodgson, Geoffrey. 2015b. "On Defining Institutions: Rules *versus* Equilibria." *Journal of Institutional Economics* 11 (3): 497–505.

Hodgson, Geoffrey. 2017. "Introduction to the Douglass C. North Memorial Issue." *Journal of Institutional Economics* 13 (1): 1–23.

Hodgson, Geoffrey. 2019. "Taxonomic Definitions in Social Science, with Firms, Markets and Institutions as Case Studies." *Journal of Institutional Economics* 15 (2): 207–33.

Howard, Lise and Alexandra Stark. 2018. "How Civil Wars End: The International System, Norms, and the Role of External Actors." *International Security* 42 (3): 127–71.

Huebner, Bryce. 2014. *Macrocognition: A Theory of Distributed Minds and Collective Intentionality*. New York: Oxford University Press.

Isaacs, Tracy. 2011. *Moral Responsibility in Collective Contexts*. New York: Oxford University Press.

Jusup, Marko, Petter Holme, Kiyoshi Kanazawa, Misako Takayasu, Ivan Romic, Zhen Wang, Suncana Gecek, Tomislav Lipic, Boris Podobnik, Lin Wang, Wei Luo, Tin Klanjscek, Jingfang Fan, Stefano Boccaletti, Matjaz Perc. 2022. "Social Physics." *Physics Reports* 948: 1-148.

Kaidesoja, Tuukka. 2013. "Overcoming the Biases of Microfoundationalism: Social Mechanisms and Collective Agents." *Philosophy of the Social Sciences* 43 (3): 301-22.

Kincaid, Harold. 2000. “Global Arguments and Local Realism about the Social Sciences.” *Philosophy of Science* 67 (S3): S667-78.

Kincaid, Harold. 2018. “Scientific Realism and Economics.” In *The Routledge Handbook of Scientific Realism*, ed. Juha Saatsi, 369-79. London and New York: Routledge.

Kincaid, Harold. 2019. “Beyond Causalism and Acausalism.” In *Explanation in Action Theory and Historiography*, ed. Gunnar Schumann, 179-93. New York and London: Routledge.

Kutz, Christopher. 2000. *Complicity: Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press.

Lebow, Richard. 1994. “The Long Peace, the End of the Cold War, and the Failure of Realism.” *International Organization* 48 (2): 249-77.

Levendusky, Matthew. 2018. Review: *Polarized: Making Sense of a Divided America*. By James E. Campbell. *Perspectives on Politics* 16 (2): 530-2.

Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

List, Christian. 2018. “What is it Like to be a Group Agent?” *Noûs* 52 (2): 295–319.

List, Christian. 2019. *Why Free Will Is Real*. Cambridge, MA and London: Harvard University Press.

List, Christian. 2021. “Group Agency and Artificial Intelligence.” *Philosophy & Technology* 34 (4): 1213-42.

List, Christian and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.

Little, Daniel. 2020. “Organizations as Actors: Microfoundations of Organizational Intentionality.” *Philosophy of the Social Sciences* 50 (3): 260–79.

Ludwig, Kirk. 2016. *From Individual to Plural Agency*. Oxford: Oxford University Press.

Ludwig, Kirk. 2017. *From Plural to Institutional Agency*. Oxford: Oxford University Press.

McCarty, Nolan and Eric Schickler. 2018. “On the Theory of Parties.” *Annual Review of Political Science* 21: 175-93.

Miller, Seumas. 2019. “Social Institutions.” In *The Stanford Encyclopedia of Philosophy* (Summer 2019), ed. Edward Zalta. https://plato.stanford.edu/archives/sum2019/entries/socialinstitutions/

Moen, Lars. 2023. “Eliminating Group Agency.” *Economics & Philosophy* 39 (1): 43-66.

Moravcsik, Andrew. 1997. “Taking Preferences Seriously: A Liberal Theory of International Politics.” *International Organization* 51 (4): 513-53.

Moravcsik, Andrew. 1998. *The Choice for Europe: Social Purpose and State Power from Messina to Maastricht.* London: Routledge.

Moravcsik, Andrew 2018. “Preferences, Power and Institutions in 21st-century Europe.” *Journal of Common Market Studies* 56 (7): 1648-74.

North, Douglass. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge: Cambridge University Press.

North, Douglass, John Wallis and Barry Weingast. 2009. *Violence and Social Orders*. Cambridge: Cambridge University Press.

Northcott, Robert. 2015. "Opinion Polling and Election Predictions." *Philosophy of Science* 82 (5): 1260-71.

Northcott, Robert. 2017. "When are Purely Predictive Models Best?" *Disputatio* 9 (47): 631-56.

Pierson, Paul and Eric Schickler. 2020. "Madison's Constitution Under Stress: A Developmental Analysis of Political Polarization." *Annual Review of Political Science* 23: 37-58.

Pinder, Mark. 2021. "Conceptual Engineering, Metasemantic Externalism and Speaker-Meaning." *Mind* 130 (517): 141–63.

Roth, Abraham. 2017. "Shared Agency." In *The Stanford Encyclopedia of Philosophy* (Summer 2017), ed. Edward Zalta. https://plato.stanford.edu/archives/sum2017/entries/sharedagency/

Rovane, Carol. 1998. *The Bounds of Agency.* Princeton: Princeton University Press.

Salam, Gavin, Lian-Tao Wang and Giulia Zanderighi. 2022. "The Higgs Boson Turns Ten." *Nature* 607: 41-7.

Satz, Debra and John Ferejohn. 1994. "Rational Choice and Social Theory." *Journal of Philosophy* 91 (2): 71-87.

Saunders, Daniel. 2020. "Optimism for Naturalized Social Metaphysics: A Reply to Hawley." *Philosophy of the Social Sciences* 50 (2): 138–60.

Schattschneider, E. E. 1935. *Politics, Pressure and the Tariff*. New York: Prentice-Hall.

Schattschneider, E. E. 1942. *Party Government*. New York: Holt, Reinhart, and Winston.

Schattschneider, E. E. 1960. *The Semi-Sovereign People*. New York: Holt, Reinhart, and Winston.

Schweikard, David. 2022. "The Agential Profiles of Organizations." *Management Revue* 33 (Special Edition): 122-39.

Schweikard, David and Hans Bernhard Schmid. 2021. "Collective Intentionality." In *The Stanford Encyclopedia of Philosophy* (Fall 2021), ed. Edward Zalta. https://plato.stanford.edu/archives/fall2021/entries/collective-intentionality/

Shapiro, Scott. 2014. "Massively Shared Agency." In *Rational and Social Agency: The Philosophy of Michael Bratman*, ed. Manuel Vargas and Gideon Yaffe, 257-93. New York: Oxford University Press.

Strohmaier, David. 2020a. "Organisations as Computing Systems." *Journal of Social Ontology* 6 (2): 211–36.

Strohmaier, David. 2020b. "Two Theories of Group Agency." *Philosophical Studies* 177 (7): 1901–18.

Tollefsen, Deborah. 2002. "Organizations as True Believers." *Journal of Social Philosophy* 33 (3): 395–410.

Tollefsen, Deborah. 2018. "Collective Intentionality and Methodology in the Social Sciences." In *The Routledge Handbook of Collective Intentionality*, ed. Marija Jankovic and Kirk Ludwig, 389-401. London and New York: Routledge.

Tuomela, Raimo. 2013. *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press.

van Hateren, Hans. 2022. "Minimal Agency." In *The Routledge Handbook of Philosophy of Agency*, ed. Luca Ferrero, 91-100. London and New York: Routledge.

Waltz, Kenneth. 1979. *Theory of International Politics*. New York: McGraw-Hill.

Waltz, Kenneth. 2000. "Structural Realism after the Cold War." *International Security* 25 (1): 5-41.

Wendt, Alexander. 2004. "The State as Person in International Theory." *Review of International Studies* 30 (2): 289-316.

Wohlforth, William. 1994. "Realism and the End of the Cold War." *International Security* 19 (3): 91-129.

Essay 3: What States Can and Cannot Do

Abstract

"State capacity" – the things a state can do – has been the subject of much debate in contemporary social science. Scholars have used it to explain a wide range of different phenomena, from economic growth to genocide, and they have also tried to explain what causes state capacity. Meanwhile, some normative political theorists have examined whether the achievement of well-functioning states is a key step on the road to justice. However, there have been few attempts to investigate what state capacity *is*. This article fills this gap by arguing that state capacity can be better understood if one draws upon the philosophical literature on agency and capacities. It argues that state capacity is not reducible to power and provides a framework for selecting a definition of state capacity, according to how one understands the state.

1. Introduction

State capacity has been an extremely influential notion in contemporary empirical social science. It has been used to explain a wide range of phenomena such as economic growth, corruption, democratization, interstate and civil wars, genocide and it also has been itself an *explanandum*, with much research trying to identify what causes state capacity (Soifer 2008; Soifer and vom Hau 2008; Lindvall and Teorell 2016; Berwick

and Christia 2018). While less attention has been paid to it in normative political theory, some political theorists have been interested in the question of whether a well-functioning state is a precondition for the pursuit of justice (Pettit 2023).

Despite all this interest in the topic, with rare exceptions (most notably Lindvall and Teorell 2016), there have been few systematic and theoretically rigorous attempts to analyze what state capacity *is*. This article seeks to fill this gap in two ways. First, it clarifies the relation between state power and capacity. Second, it discusses what state capacity is by drawing on the philosophical literature on the nature of capacities in general.

The paper will be structured as follows. Section 2 examines the nature of the state; section 3 considers what agents are; section 4 turns to the question of what power is; section 5 argues that state capacity should not be conflated with state power; section 6 explores the nature of capacities; section 7 provides an analysis of state capacity; and, finally, section 8 concludes.

2. What is the state?

The expression "the state" means different things to different people. But in this article, I will use the definition of "the state" provided by Christopher Morris, which, I think, captures how most political scientists use the term. According to him, *the state* (or, more specifically, the modern state) is a centralized, coordinated and hierarchical form of political organization in a delimited territory that endures through time regardless of changes in leadership, is distinct from its ruling class and population, is an agent in its

own right but acts through its institutions (especially the government, the judiciary, the bureaucracy, and the armed forces), is the ultimate (*de facto*) political authority and claims a monopoly of the legitimate use of force in its territory (Morris 1998, 45-6).

Although the expression "the state" has different meanings, disputes over the nature of the state are not merely verbal. Morris' definition of "the state" picks out the so-called "neo-Weberian"[1] account of political organization, which relies on two substantive philosophical assumptions, which researchers interested in state capacity should bear in mind.

First, it assumes that methodological holism or ontological holism (or both) are true. *Methodological holism* is the view that social science can accept both individualist explanations, which refer only to the behavior of individuals (or the aggregate of individual behaviors), and holist explanations, which accept as their *explanans* both the behavior of individuals and the workings of social entities such as social class, race, the state, and social norms. This stands in contrast to methodological individualism, which only admits individualist explanations (Zahle 2021). *Ontological holism* is a view not about the nature of explanation, but about the nature of reality. It holds that social entities exist, and these are composed of or supervene upon human beings but are not a mere aggregate of them. They exist in their own right, so to speak. The opposite view is ontological individualism, which holds that words such as "the state" or "class" are just a useful shorthand but do not pick out entities that exist over and above individual human beings (Zahle 2021).

[1] I borrow the term from vom Hau (2015, 135).

Second, the neo-Weberian analysis of political organization assumes that groups of people can be intentional agents and that states are a kind of intentional group agent. There is a vast body of literature within philosophy examining whether groups can be agents (Roth 2017; Schweikard and Schmid 2021) and whether states, in particular, are agents (Lawford-Smith 2019: ch. 4). So by asserting that states are agents, the neo-Weberian is making quite a significant claim.

The two most influential alternative approaches to political organization - liberalism and Marxism - deny one or both philosophical assumptions. (Note that I am *not* talking here about liberalism and Marxism as *normative* political theories.) Liberalism rejects both the methodological/ontological holism and the realism about intentional group agency of the neo-Weberian, whereas Marxism accepts holism but denies that states are intentional group agents in their own right. This is reflected in how liberals and Marxists define "the state." Liberalism holds that "[s]tates are sites of strategic interaction among powerful individuals and groups" (vom Hau 2015, 134) while Marxism maintains that states are arenas where members of different social classes interact (vom Hau 2015, 133). In other words, neither liberalism nor Marxism accepts that states are intentional agents.

This has far-reaching implications for how one understands state capacity. If one is a Marxist or a liberal about the state, then one should not believe that states are able to perform actions. After all, it is only *intentional agents* that are capable of that (Piñeros Glasscock and Tenenbaum 2023, §2). Nevertheless, those who deny the intentional agency of states can still acknowledge the reality of one kind of state capacity, namely the ability to produce mere behavior (i.e. non-intentional action). Failing that, they

would either have to adopt (a) an *eliminativist approach* to state capacity, according to which talk of "state capacity" is just a manner of speaking, a metaphor, and does not pick out any distinctive phenomenon in the world or (b) a *stipulative approach* to state capacity, according to which the expression "state capacity" is whatever a researcher chooses it to mean, for whatever research purpose they have.

3. What is an agent?

An agent is a system that possesses at least the following three properties (Barandiaran, Di Paolo and Rohde 2009, 369-73). First, it is a coherent whole, distinct from the environment in which it is situated. Second, it is the source of at least some of its interactions with the environment. In other words, it is not like a football that only moves around if it is kicked by a person or blown by the wind. Third, the system's interactions with the environment can either be a success or a failure or somewhere in-between.

Systems that have these properties are quite simple agents. Some scholars have called these "minimal" or "primitive" agents (Burge 2009, 256-78; 2010, 327-41; van Hateren 2022). However, there are more sophisticated forms of agency than minimal agency. For example, there is *perceptual* agency – the kind of agency possessed by systems that can perceive (e.g. see or hear) the external world; and there is *intentional* agency – the kind of agency possessed by most human beings, which involves the ability to form beliefs, intentions and desires.

Whether states are minimal agents, intentional agents or not agents at all has important implications for how we understand state capacity, as I have already noted (I will return to this point later on).

4. What is power?

The word "power" is ambiguous and has at least four major meanings: a metaphysical-Aristotelian, a socio-political, a capacity and an actualist sense.

First, in the *metaphysical-Aristotelian* sense, "power" is just another word for "disposition" or "potentiality" and denotes the manifestation of a property by objects or individuals under certain conditions (Choi and Fara 2018, §1). For example, a fragile glass has the "power" to break, sugar has the "power" to dissolve in water and a short-tempered individual has the "power" to lose his mind easily (Vetter 2015, 11). This notion of "power" was introduced by Aristotle and is the basis for one prominent approach to metaphysics (i.e. the philosophical study of the nature of reality at a fundamental and general level).

Second, in the *socio-political* sense, "power" denotes a relation between two or more agents (or sets of agents). Here there are at least two possible interpretations – a *modal* and an *actualist* interpretation. According to the *modal* interpretation of socio-political power, power is the capacity of an agent to bring about an outcome that impacts on another agent in some significant way. An agent who has power in this sense possesses it even if she never exercises it. For example, an armed person robbing a bank has the power to coerce the bank's staff into helping her load her van with stolen cash

even if she does not end up demanding assistance and loads everything herself. The modal interpretation is the most natural way of interpreting both Max Weber's and Robert Dahl's famous definitions of power. Weber held that "[p]ower can be defined as every *Chance*, within a social relationship, of enforcing one's own will even against resistance" (Weber 2019 [1921], 134). Similarly, Dahl's (1957, 202-3) formulation states that "*A* has power over *B* to the extent that he can get *B* to do something that *B* would not otherwise do." By contrast, according to the *actualist* interpretation, power is the *actual* bringing about of an effect by an agent on another agent. This is how Dahl himself interpreted his own definition of power (Dahl 1968, 410; Morriss 2002 [1987], 15) and it was also how the early Steven Lukes (1974) and Michel Foucault (1983, 217) understood "power." For example, Lukes (1974, 27) argued that power is the actual instantiation of a relation according to which "A exercises power over B when A affects B in a manner contrary to B's interests." (Note, though, that later on he abandoned the actualist interpretation of socio-political power in favor of the modal interpretation. He claimed then that "power is a *dispositional* concept, identifying an ability or capacity, which may or may not be exercised" [Lukes 2005, 109; my emphasis]. For discussion of this change in Lukes' perspective, see Morriss 2006).

Third, in the *capacity* sense, "power" is just a synonym for "capacity" or "ability" and denotes both those capacities that involve relations between agents and those that do not. This definition goes back to at least Thomas Hobbes' *Leviathan* (Goldman 1972, 221; Abizadeh 2023, 3). Hannah Arendt was a notable supporter of power-as-capacity. Although, to my knowledge, she did not define capacity in general,

she used the word "power" to refer to *collective* capacities (Arendt 1970, 44).[2] In analytic philosophy (the dominant philosophical tradition in the English-speaking West), Alvin Goldman was an early advocate of the definition of power as capacity. He defined "power" (or "individual power") vis-à-vis specific issues as the capacity to bring about the outcome that one wants upon performing certain actions (Goldman 1972, 226).[3] Years later, Peter Morriss (2002 [1987]) also defined power as capacity, though, unlike Goldman, he argued that capacities are a kind of disposition.

Fourth, in the *actualist* sense, "power" just means "the production of intended effects" (Russell 2004 [1938], 23). Like the metaphysical-Aristotelian and capacity senses, the focus here is not simply on relations between individuals, but on a much more general notion. However, unlike those interpretations, "power" in the actualist sense only refers to *actual*, as opposed to possible events.

[2] For early discussions, see Lukes (1974, 28-31) and Habermas (1977). For more recent discussion see, e.g. Parietti (2022); Abizadeh (2023); Kolodny (2023). Many philosophers have examined collective capacities, though they have differed in what they call these. Goldman (1972, §2), for example, preferred the expression "collective power" while Abizadeh (2023, 4-5, 14-5) chose "agential social power." Outside the literature on "power," many theorists working on the moral duties and responsibilities of groups do not use the word "power" at all to refer to collective capacities, preferring instead less ambiguous terms, such as "collective abilities," "joint abilities," "group abilities" or "plural abilities" (e.g. Lawford-Smith 2012, 463; Pinkert 2014, 194; Aas 2015, 15; Collins 2019, 71; forthcoming, §1; Estlund 2020, 246; Miller 2020, 200; Wringe 2020; Schwenkenbecher 2021, 54).

[3] He distinguished this from "overall power," which he defined as including additional elements such as power over people and power over more than one issue Goldman (1972, 260).

5. Why state capacity is not power

State capacity is a *capacity* or *ability* and should be treated as such. (I use "capacity" and "ability" interchangeably.) It is not power, if we understand "power" in the metaphysical-Aristotelian, socio-political or actualist senses of the word.

5.1. State capacity is not power in the metaphysical-Aristotelian sense

It is a mistake to conflate capacity with metaphysical-Aristotelian power (i.e. disposition/potentiality), as Morriss (2002 [1987], ch. 4) and Lukes (2005, 109) have done.

First of all, the Aristotelian approach to metaphysics, in which dispositions play a central role in describing and explaining the fundamental nature of reality, is just *one* way of understanding the world and it is not even the dominant one among philosophers (Williams 2019, ch. 2), though it has witnessed a renewed surge of support in recent decades (Bird 2016, 341; Hansson Wahlberg 2020, 1357-8). So why should anyone who is not a staunch Aristotelian accept that capacities are dispositions? Why not adopt the dominant metaphysical position in philosophy, neo-Humeanism, and the associated analysis of capacities as restricted possibilities, which holds that "an agent has an ability to ϕ if and only if it is possible, in a properly restricted sense, for the agent to ϕ" (Jaster 2020, 63)?[4] After all, if social scientists (including social and political theorists) believe

[4] Of course, not everyone who subscribes to neo-Humeanism adopts this analysis of abilities, but there is certainly a tendency among neo-Humeans to adopt the restricted possibility analysis of abilities or something similar to it.

in the division of epistemic labor, and if neo-Humeanism is the default view among metaphysicians, then it seems reasonable for the social scientist to defer to the dominant position (assuming, of course, that one accepts the legitimacy of metaphysics as a field of inquiry).

Second, in any case, even some prominent *Aristotelian* metaphysicians reject the idea that capacities are dispositions. Barbara Vetter has argued, for instance, both independently and in joint work with Romy Jaster,[5] that the analysis of abilities as a kind of disposition is unconvincing (Vetter and Jaster 2017; Vetter 2019; see also Huoranszki 2022, 40-8). For simplicity, let us focus on one version of the dispositionalist account of abilities, the "new dispositionalism". According to it, "*[a]n agent has the ability to* Φ *[if and only if] she has the disposition to* Φ *when she tries (intends, chooses, or wants) to* Φ" (Vetter and Jaster 2017, 4; see also Fara 2008: 848). There are two major problems with this analysis.

First, there is the "problem of averted attempts." Imagine an able-bodied individual who is in a coma. If she tried (intended, chose, wanted) to raise her arm, she would have the disposition to raise her arm. Therefore, the conditions required by the

[5] In later work, Jaster (2020, 163-8) argues that "general agentive abilities" (e.g. the ability of a swimmer to swim even if there is nowhere to swim nearby) and "general non-agentive abilities" (e.g. the ability to digest food) are kinds of dispositions, but "specific abilities" (e.g. the ability to swim in Lake Zurich on Christmas day, 2022) are not. We do not need to examine her argument here. Suffice it to say (I will return to this point below) that the abilities that matter for most empirical social science and normative political theory are *specific* abilities. So even if Jaster (2020) is right, it remains the case that *state* capacity, at least, is not a disposition.

new dispositionalist account of abilities would be met. Yet it would be odd to ascribe the ability to raise her arm to her because she is in a coma. So there is a disconnect here. Although solutions have been offered to address this challenge, the jury is still out (Vetter and Jaster 2017, 4-6).

Second, there is the "problem of reliability," which is the issue that "[a]gents may have abilities that they are not disposed to exercise, even when trying to do so" (Vetter and Jaster 2017, 6). For example, "[t]he ability to write great poems, to compose beautiful music, or to produce innovative art may not be reliably triggered by the artist's trying to do these things. In fact, trying may be counter-productive. What is needed may, rather, be inspiration, a spur of the moment; the artist may feel 'overcome' with an idea over lunch, or while taking a shower, after days, months, or years of unproductive trying. In such cases, it does not seem to be the case that the artist is disposed to produce the creative feat upon trying to do so" (Vetter and Jaster 2017, 6; see also Vetter 2019). Another example would be Usain Bolt at the height of his career as a sprinter. He had the ability to run 100 meters in 9.58 seconds but he did not have the disposition to do so when he tried, given that he usually did not succeed in exercising that ability (Vetter and Jaster 2017, 6).

Whether one believes or not that capacities are dispositions is not a pointless esoteric dispute. It can affect how one conducts theoretical and empirical research. If one is a dispositionalist, who holds that for an agent A to be able to effect an outcome X on an agent B is just for A to be disposed to produce an outcome X on B upon A's trying to do so, then it seems clear that one should prioritize in one's research the dispositions that A has and not the particular relation of A having an effect on B. If

dispositions are *intrinsic* properties, as many philosophers believe they are (Choi and Fara 2018: §5), a great deal of the information about the disposition A has over B could be acquired by looking at the nature of A on its own. For example, in empirical research on the causes of the Rwandan genocide, a researcher would likely focus on the intrinsic properties of the Rwandan state, instead of on the relations between the state and the non-state agents who were essential to the perpetration of the atrocities.

If dispositions can be both *intrinsic* and *extrinsic* (as argued by, for example, Huoranszki 2022: ch. 3; McKitrick 2003; Shoemaker 1980; and Vetter 2013), a dispositionalist account of abilities would still issue different guidance to empirical researchers compared to, say, a restricted possibility analysis of abilities. For example, a study investigating how a country's capacity to derive income from natural resources affects levels of state corruption, would, if taking for granted a dispositionalist perspective, prioritize certain things over others. It would look at the country's intrinsic properties (e.g. the property of having large amounts of oil) and its extrinsic properties (e.g. the property of state officials' being unresponsive to citizens' demands, say, because they do not rely on income from taxation) and try to identify causal mechanisms that could explain the effect of natural resources capacity on corruption. By contrast, someone committed to a restricted possibility analysis of abilities might prioritize counterfactual analyses that use computer simulations such as agent-based modeling and are based on some modal logic, instead of focusing on the identification of causal mechanisms. That is, he or she might be content with a more idealized model of the effects of natural resources capacity on corruption, one that makes a number of counterfactual assumptions.

I am not saying here that states do not have dispositions. They might well do. Perhaps certain states are disposed to descend into civil war while others might be disposed to experience long periods of economic stagnation. To establish the truth of these claims, though, requires detailed theoretical arguments and empirical research and, in particular, the rejection of alternative explanations that explain the occurrence of civil wars and economic stagnation by reference to something other than a state's dispositions. Whatever the case, a state's dispositions are distinct from its capacities, so that demonstrating the existence of the former does not entail demonstrating the existence of the latter.

5.2. State capacity is not power in the socio-political sense

As Barry (1988, 341) points out, "whereas all power is ability, not all ability is power." A rich person's ability to eat caviar and drink champagne is just that, an ability. It does not entail the possession of socio-political power - although, obviously, the same wealth that buys expensive food can also be used to buy socio-political power (Barry 1988, 341).

The same principle extends to state capacity. States can do more than impact the behavior or welfare of its citizens in a significant way, so not all state capacities are cases of socio-political power, contrary to what Lindvall and Teorell (2016) argue. For example, a state can require that every office in the state bureaucracy display a picture of the head of state on one of its walls, but that is not an instance of socio-political power, given that the bureaucracy is part of the state and those outside the bureaucracy

are unaffected. Another example is that if all the ordinary citizens of a state suddenly died and only members of the government, judiciary, armed forces and so on survived, the state would still retain its capacity to enforce law and order, even if there are no ordinary citizens around to break the law.

To be sure, many interesting cases of state capacity are also cases of socio-political power, such as the capacity to enforce laws via physical coercion. However, for many research purposes, there is no need to appeal to socio-political power in addition to state capacity. Let us assume, for illustrative purposes, that Weber's analysis of socio-political power as the ability to implement an outcome despite resistance is correct. Consider comparative historical analyses of how state capacity impacted economic growth throughout the centuries. If we want to examine how an impersonal bureaucracy or standardized compulsory public schooling contributed to economic growth, it seems sufficient to focus on the actual existence of these state capacities, instead of also focusing on whether there might have been *resistance* to the establishment of an impersonal bureaucracy or standardized education in a counterfactual scenario.

Another example, now from normative political theory, would be the following. One could argue that the demands that citizens can rightfully make upon a state should be constrained to those outcomes that the state can deliver, instead of idealistic ones that it cannot. For instance, citizens of Nigeria might reasonably be entitled to demand a substantial decline in corruption, but they cannot demand that the state should guarantee that everyone becomes a billionaire. However, it is not clear, without further argument, why moral requirements should be constrained by what states can do *despite*

resistance, instead of what they can do *tout court*. Whatever position one takes on this debate, it is clear that state capacity is the more fundamental notion upon which socio-political power is based.

5.3. State capacity is not power in the actualist sense

There is more to capacity than "the production of intended effects." It is commonly believed that if an agent intentionally did something successfully, she was able to do it, even if her success was a mere fluke and will never happen again (Jaster 2020, 117). The fact that England colonized parts of North America shows that it was able to do that. However, we also want to know what England (now the United Kingdom) can do today. In order to acquire this knowledge, one needs to understand capacity in a broader way as also involving possible, and not merely past, events.

6. What is capacity?

6.1. Two desiderata that a definition of capacities must meet

It is highly controversial what capacities are (Maier 2022). And to make things worse, the debate over the nature of capacities has been scattered across a number of different disciplines and research programs, which have had little interaction with one another. Jaster (2020, 4-5) observes that there have been two major research areas in which the nature of capacities has been investigated but which have evolved as virtually isolated traditions. First, since the publication of G. E. Moore's book *Ethics* (1912), there has

been a debate within analytic philosophy about how capacities should be understood in the context of attempts to show that free will is compatible with a deterministic world. Second, Jaster notes, there has been a research program within linguistics and within certain areas of analytic philosophy that are sympathetic to the use of formal methods, focusing on the logic and semantics of words such as "can," "must" and "may."

In addition to these two areas, capacities have also been analyzed, as we have seen, in the "power" debate and in the controversy over the moral duties and responsibilities of both coordinated and uncoordinated groups of individuals (see footnote 2), in the literature on the "capability approach" in philosophy and economics (Robeyns and Byskov 2023) and, of course, in the debate over the nature of state capacity. All of these research programs have been carried out mostly in isolation from one another.

In light of this diversity of areas in which capacities have been studied, I believe that the best way to proceed is to adopt a definition of capacity that abstracts away from the particular concerns of individual research programs. This prevents the definition from being objectionably *ad hoc*. It is tempting for scholars to choose a definition that matches their preferred substantive position in a debate and to lose sight of objections to that definition that might be clear to those theorists working in other fields. For example, think of the free will debate. Those philosophers sympathetic to the argument that free will is compatible with determinism for a long time tended to adopt a view of capacities that provided a clear solution to the challenge of vindicating compatibilism,

but that is now widely seen as fraught with shortcomings (Jaster 2020, ch. 2).[6] It is plausible that if early on in that debate, other scholars who were not invested in the free will debate had paid attention to the analysis of capacities dominant in that area, they might have spotted the problems with it and paved the way for the development of a more fruitful definition.

In the case of the debate over the nature of state capacity, providing an abstract definition of capacity also helps prevent bias in favor of certain substantive conclusions. Lindvall and Teorell (2016: 3) observe that the literature on state capacity "tends to treat the development of states in a particular *region* (Western Europe) during a particular *period* (from the sixteenth century to the present) as an explicit or implicit benchmark of high state capacity." This tendency opens the door to charges of Eurocentrism and also implies that high state capacity is necessarily a good thing that should be aimed at, which could compromise empirical research not only on normative but also on epistemic grounds.[7] Suppose, for example, that it is the case that genocides are more

[6] Note, however, that this view of abilities, known as the "simple conditional analysis," has also been defended on independent grounds unrelated to the issue of compatibilism. For example, some have argued that to the extent that the conditional analysis is concerned with settling the matter of the possibility of agents' control over their actions, the analysis can avoid the objections levelled against it in the context of the compatibilism debate (Huoranszki 2011).

[7] These are my points. Lindvall and Teorell (2016: 3) maintain, instead, that the problem with the method of deriving a definition of high state capacity based on observations about the development of modern Western European states is that "it encourages us to think teleologically ('How do you get to Denmark?') and inductively ('What have the Danes been up to?')."

likely in countries with high state capacity than in those with low capacity. If one conflates high state capacity with something normatively desirable, such as achieving the levels of prosperity and peace found in present-day Western Europe, then one runs the risk of overlooking the explanatory role that high state capacity could play as the cause of genocides, the rise of authoritarianism and so on. As Lindvall and Teorell (2016: 3) put it, scholars of state capacity should move beyond the claim that "what we need to understand is how to 'get to Denmark'" (Fukuyama 2011: 14). Instead, "the way forward for the state capacity research agenda is a different one: what we need to do is to reason theoretically about what state capacity is, and what it requires" (Lindvall and Teorell 2016: 3). And to do that, I believe, one needs an abstract and general definition of capacity that is suitable not only for empirical research on state capacity but also suitable for other research programs, such as philosophical research on free will, the capability approach, socio-political power, etc. For a definition that is abstract and general enough to be applicable in a wide range of research contexts gives us some indication that it is not biased to generate some particular substantive conclusions beforehand.

A further advantage is that it helps unify otherwise disparate fields of inquiry. Given that all the scholars involved with the capacity literature are investigating the same or closely related phenomena (namely, capacities), a common framework should be encouraged on pragmatic grounds, so that it is easier to identify original contributions to the literature and to avoid duplication of work.

6.2. The success view of abilities

A good example of an analysis of capacities that meets the two desiderata of being (a) abstract enough to avoid domain-specific bias and (b) broad enough to provide a unified framework for theoretical and empirical inquiry is the so-called "success view of abilities," proposed by Jaster (2020, chs. 4-5). First, it is domain-neutral in that it can be used in all of the research programs I outlined above, including social scientific research on state capacity.[8] Second, as I will argue later, it can accommodate both views of state capacity that subscribe to a neo-Weberian view of the state as well as those that do not. Third, it is relatively metaphysically neutral in that it is compatible with various positions about the fundamental nature of reality, including Aristotelianism and neo-Humeanism (Jaster 2020, 8). Thus, in this article, I will adopt this analysis of capacities. I will outline its main elements in this section, leaving its implications for the definition of state capacity to the next one.

According to the *success view of abilities*, "[a]n agent has an ability […] to ϕ (where ϕ can be an action or a mere behavior) if and only if S ϕ's in a sufficient proportion of the relevant possible situations in which some S-trigger for ϕ-ing is present" (Jaster 2020, 160). "An S-trigger for ϕ-ing […] is any trigger such that ϕ-ing in response to that trigger counts as a success" (Jaster 2020, 156). In the case of actions, the relevant triggers are the intentions of the agent whereas in the case of mere behaviors

[8] Though, admittedly, it cannot be applied to *all* research programs. For example, it cannot be applied to the analysis of purely physical abilities, such as the abilities of subatomic particles. I thank Ferenc Huoranszki for pointing this out.

the relevant triggers are certain stimuli, such as sound or light (Jaster 2020, 154-5). Let me illustrate. First, consider an ability to perform an action – the ability to sing. Sarah has the ability to sing if and only if Sarah sings in a sufficient proportion of the relevant possible situations in which she intends to sing. Now, consider the ability to produce a mere behavior – the ability to hear a sound. Sebastian has the ability to hear an explosion next to his office if and only if he hears an explosion in a sufficient proportion of the relevant possible situations in which there is an explosion next to his office.

Abilities to perform actions are sometimes called *agentive abilities* while abilities to produce mere behavior are sometimes called *non-agentive abilities* (Vetter and Jaster 2017, 6; Jaster 2020, chs. 4-5).

According to the success view, agentive abilities can be strong or weak. An agent has a *strong* agentive ability to ϕ "if and only if S ϕ's in a sufficiently high proportion of the relevant possible situations in which she intends to ϕ" (Jaster 2020, 95). This is where Sarah's ability to sing falls. By contrast, an agent has a *weak* agentive ability to ϕ if the agent ϕ's "in a sufficient proportion of the relevant possible situations in which the agent intends to ψ, where 'ψ' specifies a different action type than 'ϕ', but ϕ-ing in response to intending to ψ counts as a success" (Jaster 2020, 140). For example, consider Emma, who has the ability to say witty remarks, but who only succeeds in doing so if she does *not* intend to sound witty. If she intends to be witty, she ends up being crass. She *is* able to say witty remarks - we would not deny that - but the intention that triggers her saying something witty is the intention to say *something* and not the intention to say *something witty* (Jaster 2020, 142). Or consider Peter, who is able to play "Pour Élise" from start to finish but who always fails to play the middle part on its

own if he intends to do so. He can only play the middle if he plays the entirety of "Pour Élise." He is surely able to play the middle part, but the intention that triggers the successful performance of that action is the intention to *play "Pour Élise"* and not the intention to *play the middle part of "Pour Élise"* (Jaster 2020, 143). So what distinguishes a strong from a weak agentive ability is that the former entails that the action can be done at will or upon request, while the latter does not (Jaster 2020, 144-5).

Both agentive and non-agentive abilities come in degrees. How much of any given ability an agent has depends on both the number of possible cases in which the agent succeeds in producing the behavior and the quality of the possible behaviors. In other words, the degree of an ability is the sum of the *reliability* and the *achievement* levels of an agent. How reliability and achievement should be measured and how much weight should be put on each of them depends on the context. Reliability can matter more than achievement and vice versa or they can matter equally (Jaster 2020, 103-8, 158).

Moreover, abilities can be general or specific. *General* abilities are abilities that an agent has "in view of her stable, mostly intrinsic features." *Specific* abilities are abilities that an agent has at a particular place and time - many extrinsic features of the agent are held fixed, in addition to her intrinsic features (Jaster 2020, 112). For example, a professional swimmer who is stranded in the middle of the Sahara desert has the general ability to swim, given her skills and physical make-up, but she does not have the specific ability to swim because there are no swimming pools or lakes around.

There are many different kinds of specific abilities, which vary according to the facts of the agent's situation that are held fixed (Jaster 2020, 116-7). However, three notable types of specific abilities are particular abilities, conative abilities and opportunities (Jaster 2020, 117). *Particular abilities* are those in which *all* facts of the agent's situation are held fixed. This is the kind of ability that one is talking about when one talks about England's ability to colonize North America in the seventeenth century or the ability of a person who intentionally jumped over a high fence, which she would not normally be able to do and which she only managed to due to an unusually strong gust of wind, which helped propel her over the fence. The agents did those things; therefore, given that past facts are fixed, they *could* do them. In addition to these past-oriented cases, particular abilities also include present-oriented examples. For instance, a swimmer who lacks the motivation to swim here and now lacks the particular ability to swim because her disinclination to swim is one of the facts of the situation that are held fixed (Jaster 2020, 117-9).

Conative abilities are those in which all facts of the agent's situation are held fixed apart from the agent's motivational states, such as her desire or intention to act (Jaster 2020, 119-20). So the swimmer who is unmotivated to swim *has* the conative ability to swim, because she will swim in a sufficiently high proportion of the relevant possible situations in which she intends to swim and many of the relevant possible situations include situations in which she does intend to swim. *Opportunities* are situations in which favorable external circumstances are held fixed (Jaster 2020, 120). For example, a swimmer has the opportunity to swim if she swims in a sufficient proportion of the situations in which she intends to swim and we restrict the relevant

possible situations to those suitable for swimming, such as situations in which there is a functioning swimming pool, a clear sky (in the case of an open pool), a lack of overcrowding, etc.

Before I move on to the next section, there are two further aspects of the success view of abilities that are worth highlighting. First, intentions are not considered to be actions. Unlike attempts or choices, which are usually judged to be actions, intentions are considered to be a special kind of mental state that triggers actions. Analyses of abilities that appeal to motivational states that are actions tend to lead to an infinite regress. For example, if we hold that S can ϕ is true if and only if S will ϕ if S tries to ϕ, then we are faced with the question of what it is for S to be able to try to ϕ, and for S to be able to try to try to ϕ, and so forth. By contrast, an analysis of abilities that appeals to intentions, according to Jaster, avoids this infinite regress because intentions are not actions: for S to intend to ϕ just is for a suitable mental state to trigger bodily responses toward the performance of an action (Jaster 2020, 101-3). Second, the success view of abilities can be expressed as the claim that an agent has the ability to ϕ if and only if the value of the fraction below is high enough (Jaster 2020, 96, 171):

$$\frac{\text{set of relevant possible trigger situations in which the agent } \phi\text{'s}}{\text{set of all possible relevant trigger situations}}$$

This means that if there are no possible relevant trigger situations, there are no abilities. Consequently, one can accommodate well-known cases in the philosophical literature in which the agent cannot intend to ϕ. For example, in the case of a coma patient, who cannot form the intention to raise her arm, one can conclude that she *cannot* raise her arm because the relevant intention situations will be zero. And one cannot divide by

zero, so the value of the fraction will not be high enough to ascribe the ability to raise her arm to her (Jaster 2020, 108-12, 170-3).[9]

7. What is state capacity?

Can the success view be extended to the case of states? It might strike some as counter-intuitive that an account designed with individual human agents in mind could be applied to states. This is a legitimate worry. Whether one can assuage it depends on what one wants from an account of state capacity. If one wants to provide a traditional causal explanation in which state capacity is the *explanans* or *explanandum*, then a mere extension of the success view to the case of states is unsatisfactory. A lot more theoretical work would be needed to detail the mechanisms through which the actions of individual agents should be aggregated and how the many millions of variables that are involved in large-scale complex social phenomena such as state behavior could be incorporated into a theory or model. This is an extremely difficult task. More realistically, one could aim for what Hayek (von Hayek 1955; 1994 [1964]) called an "explanation of the principle," which is an explanation that involves the identification of *possible* mechanisms that could explain *kinds* of events and processes, instead of the

[9] It is worth pointing out that although Jaster argues that the absence of relevant intention situations can avoid the infinite regress, she also concedes that in some cases there is no way around it. This is why she proposes the *weak* version of the analysis of agentive abilities in addition to the strong version. A separate question, which might pose a problem to the success view and which I will not attempt to tackle here, is whether intentions can be triggers at all, given that they are mental states and not some sort of event. I thank Ferenc Huoranszki for this point.

identification of actual mechanisms that explain particular events and processes (Herfeld 2018, 194-5). This is the kind of explanation that motivates, for example, many highly abstract and idealized models in economics (Herfeld 2018). If we are only concerned with providing an explanation of the principle, then a mere extension of the success view of abilities to states is justified. State behavior could be seen as mirroring individual behavior, even though, in reality, they are radically different.

So what implications follow from the success view for the case of states? If one accepts the neo-Weberian account of the state, according to which the state is an intentional agent, and one also accepts the success view of abilities, it is easy to see how state capacities should be understood. A neo-Weberian state can do *A*, in a *strong* sense, if and only if it does *A* in a sufficient proportion of the relevant possible situations in which it intends to do *A*. Meanwhile, a neo-Weberian state can do *A*, in a *weak* sense, if and only if it does *A* in a sufficient proportion of the relevant possible situations in which it intends to do *B*.[10]

[10] In Jaster's original analysis of abilities, intentions are intentions to do something *right away* (Jaster 2020, 101). This obviously does not work if we are trying to provide an account of the abilities of states, given that many things that a state can do it can only do over an extended period of time. Thus, an analysis of state capacities needs an account of intention along the lines of Michael Bratman's (1987) analysis of intentions as plans for future or temporally extended actions. I believe this is perfectly compatible with the success view of abilities. Jaster's objection to going beyond intentions to act right away is that "[i]ntentions to ϕ in the future can be given up, and it therefore does not count against an agent's ability, if an agent intends to ϕ in the future, but does not ϕ when the time comes" (2020, 101). However, intentions as plans are, by their very nature, relatively stable through time and resist

Let us consider some examples. First, let us examine cases of strong agentive capacities. The Holocaust seems to be a paradigmatic example. The German state under Hitler intended to perpetrate a genocide and succeed in doing so. It is plausible that Germany would have had that ability under a high enough proportion of counterfactual scenarios so that that capacity was not merely a fluke but something the German state could reliably do. Germany had the *particular* ability to commit a genocide, because it did it; it had the *conative* ability to do it, because it retained that ability even if it had not intended to commit a genocide; and it had the *opportunity* to carry out the Holocaust because the external conditions were favorable toward that aim, given the high levels of anti-Semitism, Hitler's rise to power, and widespread conformity to authority. Other candidates for strong agentive capacities possessed by states include the ability of states to become hosts of the Summer Olympics and the World Cup, the ability to sign treaties, the ability to start wars and the ability to determine immigration policy.

Second, consider now cases of weak agentive capacities. Many economic abilities seem to fall within this category. If orthodox macroeconomics is correct, a state can control inflation if it controls inflation in a sufficiently high proportion of situations in which it intends to give independence to its central bank to control monetary policy. A state can grow its GDP if it does so in a sufficient proportion of situations in which it intends to improve productivity levels, invest in science and technology, simplify regulations, etc. France can cause protests if it causes protests in a sufficient proportion

reconsideration in the absence of a change in circumstances (Bratman 1987). Therefore, they can be relied upon by the success view of abilities.

of situations in which it intends to raise the pension age.[11] All these cases are cases in which the state does not have direct control over the outcomes but in which we would, nonetheless, accept that it has *some* kind of ability vis-à-vis the outcomes.

Moreover, it seems clear that for most research purposes in empirical social science and normative political theory, it is *specific* abilities that are relevant, as opposed to *general* abilities. What matters, most of the time, is what states can do in a particular situation, given how the world is at a given time. States are far too connected to other states and to their own societies. If a state is able to grow its economy, it is always partly in light of certain facts about the global economy. If a state is able to consistently crack down on internal dissent, it is partly because the outside world is turning a blind eye. So there is little use for ascriptions of abilities that merely consider

[11] Shouldn't the intention to cause protests be related in some way to the intention to raise the pension age? After all, we saw that Emma has the (weak) agentive ability to say something witty only if she says something witty in a sufficient proportion of possible situations in which she *intends to say something*. The action that she intends, *saying something*, is part of the unintended consequence *saying something witty*. In my view, weak agentive abilities, on the success view, do not require that an agent's intended outcome be part of the unintended outcome. France is able to cause protests if it intends to raise the pension age because there is a very reliable link between its intention and the outcome, even if the outcome is completely unintended. So there *is* a relation between France's intention and the unintended outcome. "Success," as I understand it in the context of the success view, does not imply something positive, something that the agent would find desirable or harmless. It is merely a condition for the performance of an action. If France causes protests when it intends to raise the pension age, it is meeting the success condition for the performance of one action, namely causing protests. I thank Ferenc Huoranszki for pressing me on this point.

the stable, mostly intrinsic properties of an individual state. That is not to deny that focusing on the general abilities of states might be relevant in a narrow range of circumstances. For example, political theorists might want to know what an ideally just state looks like, so it would make sense for them to abstract away from many intrinsic and extrinsic features of actual states, which are far from exemplars of justice and so are unhelpful guides to finding out what a just state is.

Two noteworthy kinds of capacities that states can have are *despotic power* and *infrastructural capacity*, as Mann (1984) famously observed (though he called the latter "infrastructural power"). According to Mann (2012, 169-70), "*[d]espotic power* refers to the

range of actions that the ruler and his staff are empowered to attempt to implement without routine, institutionalized negotiation with civil society groups. A supreme despot, say a monarch whose claim to divinity is generally accepted (as in Egypt or China throughout much of their imperial histories) can thus attempt virtually any action without 'principled' opposition." This is the sort of power that the Queen in *Alice's Adventures in Wonderland* had: she was deemed to be entitled to shout "off with his head!" to anyone who displeased her and to get her wish fulfilled (Mann 1984, 189). For Mann, despotic power does not necessarily entail the ability to actually get people to comply with one's wishes – after all, he claims, many historical empires had a high level of despotic power but little ability to make people comply (Mann 1984, 191). Despotic power, then, is what political theorists usually call "*de facto* political authority" (or political authority in a descriptive, non-normative sense) – the ability of a ruler to be seen by its subjects as morally entitled to issue commands and impose rules

(Christiano 2020, §1). To avoid confusion, I suggest that we refer to the phenomenon that Mann calls "despotic power" by the more common term "*de facto* political authority." Nevertheless, there is another kind of capacity in the vicinity of *de facto* political authority that it would be apt to call "despotic power." According to my definition, which draws inspiration from Goldman's (1972, 257-60) account of "overall power," *despotic power* is the ability of a state to make a large number of its subjects comply with a wide range of commands or rules that are socially significant in the sense that these either affect people's behavior or welfare. So while the Soviet Union under Stalin had a high degree of despotic power, present-day Switzerland has, in comparison, much less of it.

Another major kind of state capacity is *infrastructural capacity*. According to Mann, this is the capacity of the state "to penetrate and centrally co-ordinate the activities of civil society through its own infrastructure" (1984, 190; see also p. 189). Examples of infrastructural capacity include the ability to tax income and wealth at source, the ability to engage in mass surveillance, the ability to quickly enforce compliance, and the ability to provide a source of income to large swathes of the population, either through employment or the welfare system (Mann 1984, 189). I believe that Mann is right here. He is also right to note that it is possible for a state to have a high degree of infrastructural capacity while having a low degree of *de facto* political authority and that most liberal democracies demonstrate this (Mann 1984, 189-90). For example, the United States has an extremely high degree of infrastructural capacity but, compared to many historical states, it has a relatively low degree of *de facto* political authority. Think of how a large proportion of the US population is

unwilling to allow the state to provide free and universal healthcare or to give up its right to own guns or how, more generally, the US (the state) is severely constrained by the interests of major lobby groups and the electorate.

How *de facto* political authority, despotic power (in my sense) and infrastructural capacity should be analyzed given the success view of abilities will depend on the state under consideration (Austria? Zimbabwe?), the type of ability in question (particular? conative?) and how fine-grained one is individuating that capacity (infrastructural capacity or capacity to hack phones?). For example, if one is taking infrastructural capacity as a whole – as the conjunction of the ability to tax, to engage in mass surveillance etc. – in the context of the current US, one might say that the US has a high degree of infrastructural capacity, but only in the sense of having a *weak* agentive capacity. Given polarization, divided government, and the pressures from lobby groups and voters, the US can only "penetrate and centrally co-ordinate the activities of civil society through its own infrastructure" if it does so in a sufficient proportion of situations in which it intends to do something else, such as engage in pork-barrel politics and appease popular sensibilities by telling voters what they want to hear. By contrast, if we individuate the US's infrastructural capacity more finely, by focusing on just its capacity to carry out mass surveillance, then it seems that we are talking about a *strong* agentive capacity. After all, the US intelligence establishment is notorious for succeeding in surveilling large numbers of people in a sufficient proportion of cases in which it intends to do so.

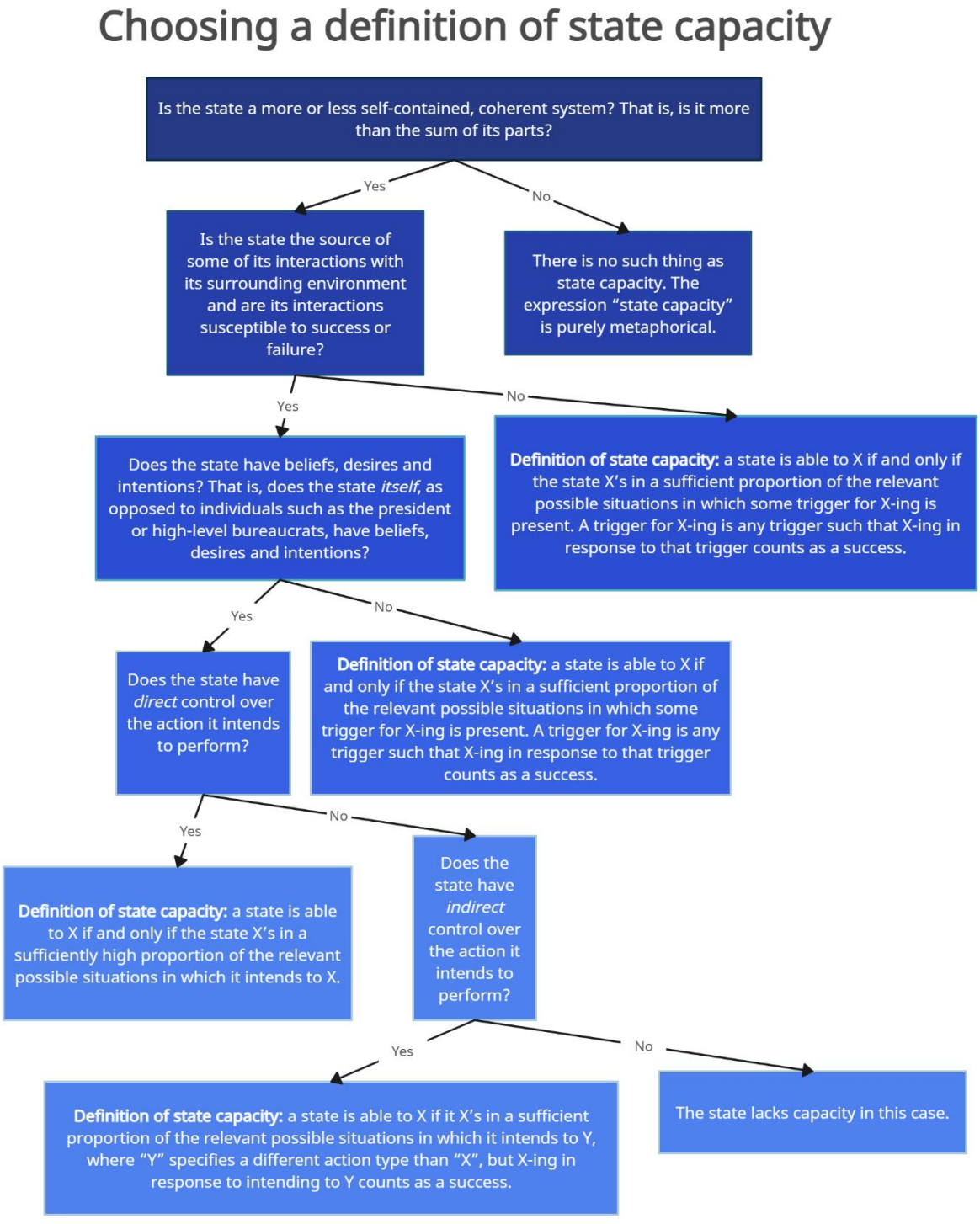
So far, I have assumed that the neo-Weberian analysis of the state as an intentional agent is right. However, what if states are not intentional agents but are,

rather, simply agents in the very minimal sense I outlined earlier? Or what if they are not agents at all? I think the success view of abilities can account for these alternative views very easily. Recall that according to it, S has an ability to ϕ "if and only if S ϕ's in a sufficient proportion of the relevant possible situations in which some S-trigger for ϕ-ing is present." If the state is a *minimal agent*, then the relevant trigger will not be an intention, but rather internal or external stimuli. For example, in the case of the ability of Nazi Germany to perpetrate a genocide, instead of analyzing that ability in terms of the intentions of the German state, we would look at the collection of triggers that led (or could have led) to the Holocaust. These would include Hitler's intentions, anti-Semitism in German society and the disposition of public officials to conform to the commands of those in authority. If a genocide would have happened in a sufficient proportion of the possible situations in which these triggers were present, then Nazi Germany had the ability to commit a genocide. So there is no need to appeal here to the intentions of the German state itself. The same principle applies if we think that the state is not an agent, but just some sort of more or less self-contained, coherent system. Whatever the system can do is a function of how it responds to internal and external stimuli in a sufficient proportion of possible situations. Although Jaster herself does not talk about non-agential systems, I see no reason why her analysis of abilities cannot be extended to them.

If the state is not even a self-contained, coherent system - that is, if it is nothing more than the sum of its parts - then there is no such thing as state capacity. Talk of "state capacity" in this scenario becomes purely metaphorical. The term will denote the ability of something or someone to do something, but the person or thing in question

will not be a state. It will be certain individuals or perhaps loosely connected groups of individuals within a given territory.

Below, I have summarized in a chart the process through which one should arrive at a definition of state capacity, depending on how one understands the state.



8. Conclusion

In this article, I have advanced a number of negative and positive theses. The negative theses were that (a) what states can do have nothing to do with what they are disposed to do; (b) state capacity is not reducible to socio-political power; and (c) it is a mistake to think that states can do only the things that they actually do. The positive theses were that (i) state capacity should be defined along the lines of the success view of abilities and (ii) it matters a great deal for how we understand state capacity whether we think that the state is a minimal agent, an intentional agent, a non-agential system or not a coherent system at all.

The aim of this article has been relatively modest. I have said nothing about how state capacity should be measured in practice in empirical research nor have I taken a stance on any arguments in normative political theory, such as the argument that state capacity is a constraint on a theory of justice. I have only tried to illuminate what state capacity is.

References

Aas, Sean. 2015. "Distributing Collective Obligation." *Journal of Ethics & Social Philosophy* 9 (3): 1-15.

Abizadeh, Arash. 2023. "The Grammar of Social Power: Power-to, Power-with, Power-despite and Power-over." *Political Studies* 71 (1): 3–19.

Arendt, Hannah. 1970. *On Violence*. San Diego: Harcourt Brace Jovanovich, Publishers.

Barandiaran, Xabier, Ezequiel Di Paolo and Marieke Rohde. 2009. "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action." *Adaptive Behavior* 17 (5): 367-86.

Barry, Brian. 1988. "The Uses of Power." *Government and Opposition* 23 (3): 340-53.

Berwick, Elissa and Fotini Christia. 2018. "State Capacity Redux: Integrating Classical and Experimental Contributions to an Enduring Debate." *Annual Review of Political Science* 21: 71-91.

Bird, Alexander. 2016. "Overpowering: How the Powers Ontology Has Overreached Itself." *Mind* 125 (498): 341-83.

Bratman, Michael. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Burge, Tyler. 2009. "Primitive Agency and Natural Norms." *Philosophy and Phenomenological Research* 79 (2): 251-78.

Burge, Tyler. 2010. *Origins of Objectivity*. New York: Oxford University Press.

Choi, Sungho and Michael Fara. 2018. "Dispositions." In *The Stanford Encyclopedia of Philosophy* (Fall 2018), ed. Edward Zalta. https://plato.stanford.edu/archives/fall2018/entries/dispositions/

Christiano, Tom. 2020. "Authority." In *The Stanford Encyclopedia of Philosophy* (Summer 2020), ed. Edward Zalta. https://plato.stanford.edu/archives/sum2020/entries/authority/

Collins, Stephanie. 2019. *Group Duties: Their Existence and Their Implications for Individuals*. Oxford: Oxford University Press.

Collins, Stephanie. Forthcoming. "Abilities and Obligations: Lessons from Non-agentive Groups." *Erkenntnis*. https://doi.org/10.1007/s10670-021-00507-5

Dahl, Robert. 1957. "The Concept of Power." *Behavioral Science* 2 (3): 201-15.

Dahl, Robert. 1968. "Power." In *International Encyclopedia of the Social Sciences*, ed. David Sills, 405-15. New York: Free Press.

Estlund, David. 2020. *Utopophobia: On the Limits (If Any) of Political Philosophy*. Princeton: Princeton University Press.

Fara, Michael. 2008. "Masked Abilities and Compatibilism." *Mind* 117 (468): 843-65.

Foucault, Michel. 1983. "The Subject and Power." In *Michel Foucault: Beyond Structuralism and Hermeneutics*, eds. Hubert Dreyfus and Paul Rabinow, 208-26. Chicago: University of Chicago Press.

Fukuyama, Francis. 2011. *The Origins of Political Order: From Prehuman Times to the French Revolution*. New York: Farrar, Straus and Giroux.

Goldman, Alvin. 1972. "Toward a Theory of Social Power." *Philosophical Studies* 23 (4): 221-68.

Habermas, Jürgen. 1977. “Hannah Arendt's Communications Concept of Power.” Trans. Thomas McCarthy. *Social Research* 44 (1): 3-24.

Hansson Wahlberg, Tobias. 2020. “Causal Powers and Social Ontology.” *Synthese* 197 (3): 1357–77.

Herfeld, Catherine. 2018. “Explaining Patterns, not Details: Reevaluating Rational Choice Models in Light of their Explananda.” *Journal of Economic Methodology* 25 (2): 179-209.

Huoranszki, Ferenc. 2011. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.

Huoranszki, Ferenc. 2022. *The Metaphysics of Contingency: A Theory of Objects' Abilities and Dispositions*. London: Bloomsbury Academic.

Jaster, Romy. 2020. *Agents' Abilities*. Berlin: De Gruyter.

Kolodny, Niko. 2023. “Partners and Patients: A Revised Grammar of Social Power.” *Political Studies* 71 (1): 20–9.

Lawford-Smith, Holly. 2012. “The Feasibility of Collectives' Actions.” *Australasian Journal of Philosophy* 90 (3): 453-67.

Lawford-Smith, Holly. 2019. *Not In Their Name: Are Citizens Culpable For Their States' Actions?* Oxford: Oxford University Press.

Lindvall, Johannes and Jan Teorell. 2016. “State Capacity as Power: A Conceptual Framework.” Working paper. https://ces.fas.harvard.edu/uploads/files/Event-Papers/State-Capacity-as-Power-September-2016.pdf

Lukes, Steven. 1974. *Power: A Radical View.* London: Macmillan Press.

Lukes, Steven. 2005. *Power: A Radical View*. 2nd edition. Basingstoke: Palgrave Macmillan.

Mann, Michael. 1984. "The Autonomous Power of the State: Its Origins, Mechanisms and Results." *European Journal of Sociology* 25 (2): 185-213.

Mann, Michael. 2012. *The Sources of Social Power: Volume 1: A History of Power from the Beginning to AD 1760*. 2nd edition. Cambridge: Cambridge University Press.

Maier, John. 2022. "Abilities." In *The Stanford Encyclopedia of Philosophy* (Fall 2022), eds. Edward Zalta and Uri Nodelman. https://plato.stanford.edu/archives/fall2022/entries/abilities/

McKitrick, Jennifer. 2003. "A Case for Extrinsic Dispositions." *Australasian Journal of Philosophy* 81 (2): 155-74.

Miller, Seumas. 2020. "Joint Abilities, Joint Know-how and Collective Knowledge." *Social Epistemology* 34 (3): 197-212.

Moore, G. E. 1912. *Ethics*. London: Williams & Norgate.

Morris, Christopher. 1998. *An Essay on the Modern State*. Cambridge: Cambridge University Press.

Morriss, Peter. 2002 [1987]. *Power: A Philosophical Analysis*. 2nd edition. Manchester: Manchester University Press.

Morriss, Peter. 2006. "Steven Lukes on the Concept of Power." *Political Studies Review* 4 (2): 124–35.

Parietti, Guido. 2022. *On the Concept of Power: Possibility, Necessity, Politics*. Oxford: Oxford University Press.

Pettit, Philip. 2023. *The State*. Princeton: Princeton University Press.

Piñeros Glasscock, Juan and Sergio Tenenbaum. 2023. "Action." In *The Stanford Encyclopedia of Philosophy* (Spring 2023), eds. Edward Zalta and Uri Nodelman. https://plato.stanford.edu/archives/spr2023/entries/action/

Pinkert, Felix. 2014. "What We Together Can (Be Required to) Do." *Midwest Studies in Philosophy* 38 (1): 187-202.

Robeyns, Ingrid and Morten Byskov. 2023. "The Capability Approach." In *The Stanford Encyclopedia of Philosophy* (Summer 2023), eds. Edward Zalta and Uri Nodelman. https://plato.stanford.edu/archives/sum2023/entries/capability-approach/

Roth, Abraham. 2017. "Shared Agency." In *The Stanford Encyclopedia of Philosophy* (Summer 2017), ed. Edward Zalta. https://plato.stanford.edu/archives/sum2017/entries/shared-agency/

Russell, Bertrand. 2004 [1938]. *Power: A New Social Analysis*. London: Routledge.

Schweikard, David and Hans Bernhard Schmid. 2021. "Collective Intentionality." In *The Stanford Encyclopedia of Philosophy* (Fall 2021), ed. Edward Zalta. https://plato.stanford.edu/archives/fall2021/entries/collective-intentionality/

Schwenkenbecher, Anne. 2021. *Getting Our Act Together: A Theory of Collective Moral Obligations*. London: Routledge.

Shoemaker, Sydney. 1980. “Causality and Properties.” In *Time and Cause: Essays Presented to Richard Taylor*, ed. Peter van Inwagen, 109-35. Dordrecht: Reidel.

Soifer, Hillel. 2008. “State Infrastructural Power: Approaches to Conceptualization and Measurement.” *Studies in Comparative International Development* 43 (3-4): 231–51.

Soifer, Hillel and Matthias vom Hau. 2008. “Unpacking the *Strength* of the State: The Utility of State Infrastructural Power.” *Studies in Comparative International Development* 43 (3-4): 219–30.

van Hateren, Hans. 2022. “Minimal Agency.” In *The Routledge Handbook of Philosophy of Agency*, ed. Luca Ferrero, 91-100. London and New York: Routledge.

Vetter, Barbara. 2013. “‘Can’ without Possible Worlds: Semantics for Anti-Humeans.” *Philosophers' Imprint* 13 (16): 1-27.

Vetter, Barbara. 2015. *Potentiality: From Dispositions to Modality*. Oxford: Oxford University Press.

Vetter, Barbara. 2019. “Are Abilities Dispositions?” *Synthese* 196 (1): 201-20.

Vetter, Barbara and Romy Jaster. 2017. “Dispositional Accounts of Abilities.” *Philosophy Compass* 12 (8): 1-11.

vom Hau, Matthias. 2015. “State Theory: Four Analytical Traditions.” In *The Oxford Handbook of Transformations of the State*, eds. Stephan Leibfried, Evelyne Huber, Matthew Lange, Jonah Levy and John Stephens, 131-51.

von Hayek, Friedrich. 1955. “Degrees of Explanation.” *British Journal for the Philosophy of Science* 6 (23): 209-25.

von Hayek, Friedrich. 1994 [1964]. “The Theory of Complex Phenomena.” In *Readings in the Philosophy of Social Science*, eds. Michael Martin and Lee McIntyre, 55-70. Cambridge, MA: MIT Press.

Weber, Max. 2019 [1921]. *Economy and Society*. Ed. and trans. Keith Tribe. Cambridge, MA: Harvard University Press.

Williams, Neil. 2019. *The Powers Metaphysic*. Oxford: Oxford University Press.

Wringe, Bill. 2020. “Global Obligations, Collective Capacities, and ‘Ought Implies Can.’” *Philosophical Studies* 177 (6): 1523–38.

Zahle, Julie. 2021. “Methodological Holism in the Social Sciences.” In *The Stanford Encyclopedia of Philosophy* (Winter 2021), ed. Edward Zalta. https://plato.stanford.edu/archives/win2021/entries/holism-social/

Conclusion

Let me summarize the main points of this dissertation. My overall argument has been that there is no such thing as feasibility, strictly speaking, and that most of the philosophically interesting questions concerning the nature of feasibility are really questions about the nature of agency and abilities. That is why I decided to focus two of the essays on the issue of whether organizations, such as states, firms, political parties, trade unions, universities and so on can be agents, and on the issue of what it is for states to have abilities.

I came to the conclusion that one important argument for organizational agency, namely "the argument from empirical social science," fails. This is troubling for philosophers who would like to see debates about organizational agency being relevant beyond philosophy. Not only have most social scientists, even many theory-conscious ones, ignored such philosophical discussions, but whenever their work postulates organizational agency, this postulation seems to be a merely useful fiction or the work is not empirically successful enough to warrant belief in organizational agency. I argued, though, that the argument from empirical social science could be given new impetus. We just need to look for empirical studies that treat organizational agency as a central independent variable.

Regardless, I would emphasize that empirical arguments are just one part of the story. There is still an important role left for largely *a priori* accounts of organizational agency. Questions about agency, like questions about conscience, free will, aesthetics

and normative political philosophy, cannot be settled by purely empirical means. Nonetheless, I believe it is helpful for theorists of organizational agency to keep empirical concerns in the back of their minds, if anything, as a means of guarding against too much introspection, which can lead one to lose sight of what the world is really like.

In this dissertation, I have also focused on the notion of state capacity, which is interesting not only in its own right, but is also a notion that plays an important role in contemporary empirical social science. It could perhaps also shed some light on statist normative political theories, which see the state as normatively important or, at the very least, as an unavoidable reality for the foreseeable future and so something that any political philosopher must grapple with. I argued that how one understands state capacity will depend on how one understands the state. The state can be seen as an intentional agent, a non-intentional agent, a non-agential system or not a system at all. This has far reaching implications for how one should analyze state capacity. However, I did not argue for any particular account of the state, restricting myself to showing what follows from each account in the context of the analysis of state abilities. I suggested that Romy Jaster's success view of abilities is a helpful way of illuminating what state capacity is. I did not provide a full-throated defense of it, only an indication of some of its virtues. The metaphysics of abilities is highly controversial and the success view, like any analysis of abilities, is not immune to objections. But one has to start somewhere and the success view seems to be virtuous enough to provide a (defeasible) foundation upon which an account of state capacity can be built.

Although I think the most philosophically interesting questions about the nature of feasibility concern questions of agency and abilities, there are other important issues that are worth exploring and that are not reducible to the latter. One example is the methodological debate over how to answer questions about what feasibility is. I argued in the introduction and in the first essay that the best approach is to start by disambiguating between the different meanings of terms to avoid merely verbal disputes and then engage in a philosophical analysis that draws upon a general theoretical framework, such as actualism. I criticized alternative views, such as conceptual analysis, conceptual engineering and one version of functionalist analysis, defended by Nicholas Southwood.

Another example of an interesting question about the nature of feasibility that is not reducible to questions about agency and abilities is whether feasibility is a property that exists in its own right or not and what exactly the necessary and sufficient conditions for a feasibility claim to be true or false are. I maintained, in the first essay, that there is no *sui generis* property of feasibility and that an analysis of feasibility can be given along the lines of "There is at least one individual agent, collection of individual agents or group agent A that can X."

Yet another interesting issue related to feasibility that is worth exploring is the relation between the debate in moral and political philosophy regarding whether feasibility is a constraint on a normative theory and the debate in the philosophy of language and mind concerning the role of feasibility in considerations about conceptual engineering. These discussions have occurred mostly in parallel, with little interaction with each other, so some insight might be gained by examining them together. The

literature in moral and political philosophy started much earlier and is therefore more developed, so it might have something useful to add to the debate about conceptual engineering. My guess is that what is true in the context of moral and political philosophy will also be true in the context of the philosophy of language and mind. For example, if the principle that "ought implies can" is true, it applies both to a theory about the moral requirement to change the semantic meaning of a word as it does to a theory about the moral requirement to bring about global distributive justice.

Note, however, that the connection between moral and political philosophy and the philosophy of language and mind in the context of the feasibility debate is probably quite restricted. For example, while the debate about the existence of feasibility constraints on normative theories in political philosophy is linked to wider issues about ideal and non-ideal theory, realism and moralism and so on, the same cannot be said about the feasibility issue in the philosophy of language and mind. As Cappelen and Dever (2021) argue, it makes no sense to extend the distinction between ideal and non-ideal theory as used in political philosophy to debates about language. For example, political philosophers who subscribe to this distinction and are interested in justice would first try to figure out what ideal justice looks like and then work out the details of how we can get from our current non-ideal circumstances to that ideal in a morally acceptable manner. There is nothing analogous to that in the philosophy of language. For instance, what would the distinction between an ideal and a non-ideal semantics for conditionals amount to? It just seems here to be a mistake to claim that there is any such distinction. The debate about idealization in the philosophy of language has more in

common with the long-standing controversy in the philosophy science about the subject (cf. Keiser 2023).

References

Cappelen, Herman and Josh Dever. 2021. "On the Uselessness of the Distinction Between Ideal and Non-Ideal Theory (At Least in the Philosophy of Language)." In *The Routledge Handbook of Social and Political Philosophy of Language*, eds. Justin Khoo and Rachel Sterken, 91-105. New York: Routledge.

Keiser, Jessica. 2023. *Non-Ideal Foundations of Language*. New York: Routledge.