THE UNIVERSITY OF CHICAGO

FREEDOM

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE HUMANITIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

BY

PASCAL BRIXEL

CHICAGO, ILLINOIS

AUGUST 2020

For my parents

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

In this dissertation, I develop and defend a conception of external freedom: freedom in relation to other people and the extrapersonal material world. One does something freely, I argue, only if one does it for its own sake and not merely for the sake of further ends. Historically, this idea is rooted in the work of Aristotle and especially Marx. It is more or less foreign, however, to the mainstream of contemporary analytic philosophy, and it has controversial political implications, particularly with respect to the social organization of labor.

Mainstream accounts of freedom can be divided roughly into "negative" and "positive" theories. "Negative" theories tend to construe freedom as the absence of interpersonal coercion. Such approaches, I argue, are unable to make sense not only of non-interpersonal unfreedom and its social significance but even of coercion, for they fail to explain the way in which coercion makes its victim unfree, and consequently explain the wrongfulness of coercion in the wrong way. "Positive" theories, on the other hand, tend to construe freedom as autonomy, understood in terms of practical reason or desire. I argue that in their currently dominant forms, these theories are also unable to make sense of coercion because they misconstrue the unfreedom of coerced action as a defective form of internal self-relation, such as an impaired ability to set one's own ends or to adjust one's lower-order desires on the basis of one's higher-order desires.

My own account is a "positive" conception of freedom as autonomy which makes room for genuinely external forms of unfreedom. I begin with the familiar idea

that doing something freely requires doing what one wants to do. But the relevant kind of desire, I argue, is *intrinsic* desire—the desire to do something on account of its intrinsic value. It follows that one does something freely only if one does it for its own sake, and not *merely* instrumentally. Free activity must be, in Marx's terms, not "merely a means to satisfy needs external to it" but itself "the satisfaction of a need."

I show that this view makes sense of paradigm cases of unfreedom including the unfreedom of coerced action. Coerced action, by virtue of the peculiarly external way in which it is incentivized, is essentially a species of merely instrumental action and therefore unfree. I argue, however, that other activities motivated by similarly external incentives—such as paid labor—are merely instrumental, and therefore unfree, in the same way. Notwithstanding some moral differences between them, both coercively threatening someone and incentivizing them by means of payment are ways of objectionably using that person as a mere means.

Finally, I show that toil—labor which is intrinsically unchoiceworthy by virtue of its content—is as such unfree. This includes "bullshit jobs" which the worker considers pointless as well as routine labor which affords the worker no meaningful scope for deliberation in her work. I argue that a market economy can be expected to give rise to avoidable toil of both kinds, and hence to avoidable unfreedom. The market, moreover, is also an unsuitable mechanism for the distribution of *un*avoidable toil, since such toil, as a form of socially necessary unfreedom, constitutes a distinctive social burden incommensurable with socioeconomic goods. If we are genuinely committed to living in a society of free people, I conclude, we have a collective duty to seek an alternative to the market as the central form of the social organization of productive activity.

ACKNOWLEDGMENTS

I could not have begun or finished this dissertation without the help of many friends, colleagues, and teachers. For keeping the Philosophy Department running and for their absolutely dependable administrative support on many occasions, I am grateful to William Weaver and Jessica Barbaro. For their role in preparing me for this stage of my education in the first place, I am grateful to many past teachers, especially Avner Baz, the late Bob Hargrave, Shannon Leggett, and Tony Ackerman. I am also grateful for the support and solidarity of many friends and colleagues, especially Dan Paget, Rory O'Connell, Amichai Amit, Jacob Butcher, and the Philosophy graduate students at the University of Chicago, as well as organizers and members of TPL and GSU, especially Jacob Swenson-Lengyel, Will Tanzman, and Jake Werner.

My most direct thanks go to my committee. A.J. Julius came onboard relatively late in the process, but offered generous and invaluable advice. While I always get the feeling, when I read his work, of having encountered a fellow traveler and kindred spirit, his criticisms significantly improved this dissertation. I am particularly grateful to A.J. for pressing me to take more seriously the distinctiveness of coercion, and the magnetism of the thought that an ideal of freedom must be fully attainable in principle—that "the fix must be metaphysical, not technological." I am still grappling with some of his challenges, and I look forward to many more conversations.

Dan Brudney's influence on this project extends from its overall shape down to the details. It was in one of his classes that I first read Marx's early writings, and Dan's

own work on the early Marx has also been an inspiration. The many challenges he has put to me in conversation, moreover, have been indispensable to me. To begin with, he convinced me that my conception of freedom would require a much more substantial defense than I had initially planned for. Since then, his feedback has improved every part of my dissertation and made my arguments more careful and patient on countless occasions, especially when it came to responding to alternative accounts of freedom and autonomy.

Ben Laurence has likewise helped me in fundamental ways. His objections, questions, and reading suggestions always targeted the most important issues and were unfailingly useful, especially in putting me into conversation with the Kantian and republican traditions and in pressing me to clarify the role of the concept of freedom in second-personal claims. I also want to mention that I immensely appreciated Ben's encouragement along the way, which helped to sustain me at difficult moments. More generally, both Ben's teaching and his writings in political philosophy have been formative for me. I am grateful especially for his continued reminder that we have to do political philosophy in such a way that we can say with a straight face that what we are doing is part of a real, practical endeavor of social change.

Anton Ford, who chaired the committee, has been one of my most important teachers and mentors. The clarity and systematicity of his thought sets the standard to which I aspire in philosophy. More concretely, it was in his seminar that I first read *Capital*, and Anton's understanding of Aristotle and Marx has shaped this dissertation in too many ways to count. Throughout the long process of writing, moreover, Anton has been generous with his time and provided direct, detailed, and invaluable substantive feedback. Equally importantly, he has consistently offered emotional

support and practical advice, and I am not sure I could have gotten this project off the ground if he had not believed in it from the beginning, when it was a very ill-formed and indeterminate thing. Finally, I am grateful to Anton for his role in inspiring me to become involved in real political organizing, beginning in the days of Occupy.

Most of all, I owe immense gratitude to the four people closest to me. I do not really know how to thank Claire Kirwin, because there is too much to be thankful for. So let me just say that graduate school was much harder than I ever thought it would be, but Claire made even the most difficult parts of these many years worthwhile. She has also read and commented on multiple drafts of most of this dissertation, and in fact I only chose to embark on the project in the first place thanks to her encouragement.

Thanks also to my brother, Marcel, for his companionship—and for his own dream of freedom and his inspiring refusal to believe that anything is impossible. And thanks, finally, to my parents for the long and hard labor of raising me, nurturing me, and providing for me materially, as well as for giving me the freedom to make my own way in life. They have made everything possible for me, and I dedicate this dissertation to them with love and gratitude.

INTRODUCTION

1. TWO DESCRIPTIONS OF WORKING LIFE

Many people in modern market societies spend the bulk of the waking hours of their adult lives doing work which they perform purely instrumentally, not on account of its intrinsic value but only to pay the bills. Can such societies be societies of free people?

Consider the following two descriptions of working life, which I believe are in important respects typical. Here is Phil, a spot-welder in the U.S., describing his job to Studs Terkel in the early 1970s:

> The welding gun's got a square handle, with a button on the top for high voltage and a button on the bottom for low. The first is to clamp the metal together. The second is to fuse it. […] I stand in one spot, two- or three-foot area, all night. The only time a person stops is when the line stops. We do about thirty-two jobs per car, per unit. Forty-eight units an hour, eight hours a day. Thirty-two times forty-eight times eight. Figure it out. That's how many times I push that button.[1]

The number is 12,288. He goes on to say: "I know I could find better places to work. But where could I get the money I'm making? Let's face it, $4.32 an hour. That's real good money now."[2]

Here is a second example. Virginia, interviewed by Lauren Greenfield a few years ago, is a Filipina who works as a nanny for the family of a billionaire in Florida.

---

[1] Studs Terkel, *Working: People Talk About What They Do All Day and How They Feel About What They Do* (New York: Random House, 1972), 159, superfluous quotation marks and italics removed.
[2] Ibid., 162. $4.32 in 1972 is equivalent to about $27 in 2020.

She works for the family five days a week, from when she gets up in the morning until the children go to sleep at night; the remaining two days of the week, she works a second job at a grocery store. Explaining the point of it all, she tells Greenfield:

> Every Filipino's dream is to have our own house. Those Filipinos that are here have their families in the Philippines, and they have big houses. I bought a piece of land there, so, if God permit, I can go home. Maybe I will build a house for that piece of land that I bought. I need some money to go home. If you go without money, it's not easy, so I need to save some money.[3]

At the time of the interview, Virginia has not seen her own children in 19 years.

Phil and Virginia are not enslaved, enserfed, indentured, or conscripted. They seem to have access to a number of employment opportunities with different employers. They do not appear to be living in dire poverty. Nor do they labor under the spell of irrational or compulsive desires. And yet their accounts, it seems to me, are descriptions of unfreedom. In this dissertation, I develop and defend a conception of external freedom—freedom in relation to other people and to the extrapersonal material world—which is capable of bringing this unfreedom into view.

2. FREEDOM AS ACTIVITY FOR ITS OWN SAKE

Roughly speaking, the mainstream theories of freedom in analytic philosophy fall into two families, which we could risk calling "negative" and "positive" theories. Negative theories typically conceptualize freedom purely as the absence of interpersonal coercion. On such a theory, Phil and Virginia could be unfree only to the extent that they are coerced. Positive theories typically conceptualize freedom as autonomy, understood

---

[3] Lauren Greenfield, *Generation Wealth* (London: Phaidon, 2017), 306.

either in terms of practical reason or in terms of desire. More specifically, these theories tend to conceive of autonomy, in effect, on the model of an internal self-relation (or so I will argue). On this kind of model, Phil and Virginia are unfree to the extent that their capacity to determine their own ends by means of practical reason or their capacity to form desires in light of higher-order desires is impaired.

In place of these approaches, I will argue that one does something freely only if one does it for its own sake and not merely for the sake of further ends. (This is a necessary and not a sufficient condition, so it is only a partial account of freedom. For ease of exposition, I will often refer to it without qualification as an account of freedom.) This is also a positive conception of freedom as autonomy, but it locates the problem with Phil's and Virginia's situations not in the defective formation of their ends or desires but in the fact that their final ends or intrinsic desires, for various reasons, fail to find expression in their activity. Both spend much of their lives doing not what they want to do but merely what they have to do. And this is the crux of their unfreedom.

The basic idea of freedom as activity for its own sake traces back to Aristotle. In the *Nicomachean Ethics*, Aristotle takes seriously only three candidates for the good human life: the life of pleasure, the life of politics, and the life of contemplation. What makes these the serious contenders is that each has some claim to be not merely instrumentally good but desirable for its own sake:

> If […] there is some end of the things we do, which we desire for its own sake (everything else being desired for the sake of this), and if we do not choose everything for the sake of something else (for at that rate the

> process would go on to infinity, so that our desire would be empty and vain), clearly this must be the good and the chief good.[4]

On one reading, Aristotle is here guided not just by a view of the good but by a view of freedom. For he rejects other ways of life, such as "the life of money-making," on the basis that they are "undertaken under compulsion," wealth in particular being "merely useful and for the sake of something else."[5] The compulsion in question here is not involuntariness in some inner sense, such as unawareness of what one is doing, as in sleepwalking; nor is it a matter of subjection to interpersonal coercion. It is rather the compulsion that consists in doing something merely as a means—out of mere necessity for an external end.[6]

In German philosophy of the 18th and 19th centuries, the Aristotelian conception of freedom is revived but comes into contact with a more expansive conception of the human good. Wilhelm von Humboldt, for instance, speaks of "[t]he true end of Man" as "the highest and most harmonious development of his powers to a complete and

---

[4] Aristotle, *Nicomachean Ethics*, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton: Princeton University Press, 1984), 1094a18–22.

[5] Ibid., 1096a6–8. Elsewhere, he says that wealth acquisition "may be studied by a free man, but will only be practiced from necessity," and that the occupations it encompasses are "the most servile" arts. Aristotle, *Politics*, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton: Princeton University Press, 1984), 1258b9ff. He is similarly dismissive of the life of labor aiming at the immediate sustenance and reproduction of life itself—cooking, eating, procreating, cleaning, etc. In the *Ethics* he simply ignores it, and in the *Politics* it appears as the life proper to slaves and animals: "both with their bodies administer to the needs of life." Ibid., 1254b25. The reading of Aristotle I am invoking here is defended by Hannah Arendt: see *The Human Condition*, 2nd ed. (Chicago: University of Chicago Press, 1998).

[6] The operative conception of freedom is therefore more robust than the basic idea of voluntariness with which Aristotle is concerned in Book III of the *Ethics*, which is compatible even with acting under coercion by threats.

consistent whole."[7] Given this broader ethical horizon, the idea of free activity as activity chosen for its own sake—i.e., because it is intrinsically good—becomes the idea not just of contemplation but of the full development and exercise of human powers more generally.

The most important defender of this idea is Marx, who famously says (in *Capital*, Volume 3) that "the true realm of human freedom" consists in "the development of human powers as an end in itself," and who argues (in the *Economic and Philosophical Manuscripts*) that free, unalienated labor is labor which is not "merely a *means* to satisfy needs external to it" but is itself "the satisfaction of a need."[8] Once again, we can distinguish two ideas here: the relatively formal conception of free activity as activity done for its own sake, and the relatively substantive interpretation of this conception in terms of the full development and actualization of human powers.

In this dissertation, I will be defending the more formal of these two ideas, for the most part leaving open the question of its substantive interpretation. Even understood formally, the Marxian conception of freedom is quite foreign to the mainstream of contemporary analytic philosophy. Versions of it are naturally defended

---

[7] Wilhelm von Humboldt, *The Sphere and Duties of Government*, trans. Joseph Coulthard (London: John Champan, 1854), 11.
[8] Karl Marx, *Capital: A Critique of Political Economy*, trans. David Fernbach, vol. 3 (London: Penguin, 1981), 958–959; *Economic and Philosophical Manuscripts of 1844*, trans. Martin Milligan (New York: Prometheus Books, 1988), 74. See also *Grundrisse: Foundations of the Critique of Political Economy (Rough Draft)*, trans. Martin Nicolaus (London: Penguin, 1973), 488, 711; and Karl Marx and Friedrich Engels, *The German Ideology*, 3rd ed. (Moscow: Progress Publishers, 1976), 242.

in some Marxian and Marx-inspired strands of philosophy.[9] Otherwise, though, the idea of free activity as activity for its own sake rarely features even as an object of explicit criticism. One of my aims in this dissertation, then, is to show that this historically rooted conception of freedom is attractive and defensible from a point of view squarely within analytic moral and political thought.

3. AN OUTLINE OF THE DISSERTATION

My argument spans five substantive chapters.

Chapter I, "Is Freedom One or Many?", is a methodological chapter in which I defend an account of what we should and can expect from a philosophical theory of freedom. The ultimate subject-matter of such a theory, I argue, should be not freedom of this or that specific kind—such as freedom from coercion, or freedom from unwanted psychological compulsions—but freedom *simpliciter*. Furthermore, a theory of this subject-matter should be systematic in the following sense. On the one hand, it should explain what distinguishes various species of freedom from one another. On the other hand, it should explain, in terms of a single, general principle, why they nonetheless fall under a common genus. I argue that only a theory which meets these desiderata has the explanatory, critical, and constructive power which belongs to an adequate

---

[9] See, e.g., Ernst Tugendhat, *Self-Consciousness and Self-Determination*, trans. Paul Stern (Cambridge, MA: The MIT Press, 1986); William Adams, "Aesthetics: Liberating the Senses," in *The Cambridge Companion to Marx*, ed. Terrell Carver (Cambridge: Cambridge University Press, 1991); Rahel Jaeggi, *Alienation*, trans. Frederick Neuhouser and Alan E. Smith (New York: Columbia University Press, 2014); and A. J. Julius, "Suppose We Had Produced as Humans," (Manuscript).

philosophical understanding of freedom, and I defend the possibility of such a theory against preliminary skeptical doubts.

Chapter II, "Kantianism, Moralism, and Psychologism," is a critique of the three main families of extant theories of freedom and coercion. All of these theories, I argue, fail to make sense of important species of freedom and unfreedom and their unity. In particular, they all fail to make sense of the unfreedom of coerced action and the wrongfulness of coercion. "Moralistic" theories attempt to explain what is wrong with coercion not in terms of the violation of autonomy but in terms of some other, independent wrong. These approaches, I argue, are unable to make sense not only of non-interpersonal forms of unfreedom and their social significance but even of coercion itself, for they fail to explain the way in which coercion makes its victim unfree, and consequently end up explaining the wrongfulness of coercion in the wrong way. "Kantian" and "psychologistic" theories, on the other hand, attempt to explain what is wrong with coercion on the basis of a positive conception of freedom as autonomy, understood in terms of either practical reason or desire. I argue that in their currently dominant forms, these theories are also unable to make sense of coercion, for they misconstrue the unfreedom of coerced action as a defective form of internal self-relation, such as an impaired ability to set one's own ends or to adjust one's lower-order desires on the basis of one's higher-order desires.

In Chapter III, "Freedom as Activity for Its Own Sake," I introduce and defend my own conception of freedom as activity for its own sake. Beginning with the familiar idea that doing something freely is, at a minimum, a matter of doing what one wants to do, I argue that the kind of desire relevant for autonomy is specifically intrinsic desire—the desire to do something on account of its (perceived) intrinsic value. It follows that

one does something freely only if one does it for its own sake, and not *merely* instrumentally. Merely instrumental activity is as such unfree. I show that its unfreedom comes in degrees; it is especially egregious when the activity is performed for the sake of obtaining basic necessities.

In Chapter IV, "Coercion and Exploitation," I develop this conception of freedom further to give an account of coercion. The basis of this account is a key distinction *within* the realm of merely instrumental action. In general, when one does something merely instrumentally, the end for the sake of which one acts is strictly distinct from the action itself. But in what is in some sense the "normal" case, the end at which one aims, though strictly distinct from the action, is nonetheless a *characteristic* product of the relevant action, and so internally connected with it through the action concept. For instance, when one undergoes transplant surgery in order to donate one's kidney to a relative dying of kidney disease, one acts merely instrumentally: no one chooses to undergo kidney surgery for its own sake. However, the end at which one aims—the health and life of the organ recipient—"belongs" to the action of undergoing kidney surgery in the sense that this is just what the relevant action is for; it is the characteristic end which gives the surgery its identity.

But it is also possible to do something merely instrumentally in such a way that one's further end in acting has nothing at all to do with the action. For instance, if one undergoes the same surgery not to save one's relative but at gunpoint, then one's further end in acting—saving oneself from being shot—is not a good which kidney transplant surgery characteristically produces. It is, as I will put it, a *radically extrinsic* end. I show that this radically extrinsic form of motivation can only be imposed on an

agent by a person and not by impersonal circumstances, and I show that coercive threats in particular essentially motivate their victims in a radically extrinsic way. But I argue that an action motivated in this radically extrinsic way is necessarily done *merely* instrumentally and therefore unfreely. In this way, I explain both why coerced action is unfree and what sets its unfreedom apart from non-interpersonal sorts of unfreedom, which involve merely instrumental motivation (as in the kidney surgery to save a dying relative) but not *radically* extrinsic forms of motivation (as in the coerced surgery). This also explains the distinctive wrongfulness of coercion.

Certain kinds of offer constitute radically extrinsic incentives in the same way as coercive threats. My account of coercion therefore implies that a certain kind of exchange, which I call "pure exchange," is unfree in the same way as coerced action, and morally objectionable in fundamentally the same way as coercion. I defend this implication, bringing both coercion and pure exchange under the common genus of "using someone as a mere means," which I also identify with "exploiting someone" and "wrongly taking advantage of someone's vulnerability." That said, I point out the possibility of other forms of exchange and cooperation which are not problematic in this way, and I show that in any case there remains a moral difference between coercion and pure exchange, which is a function of the scope of the exploited party's vulnerability.

In Chapter V, "Toil," I argue that toil—labor which is intrinsically unchoiceworthy by virtue of its content—is as such unfree. This includes "bullshit jobs"—jobs which the people doing them typically consider useless or positively harmful—and routine labor—work which affords the worker no meaningful scope for deliberation in her work. I argue that a market economy can be expected to give rise to

avoidable toil of both kinds, and hence to avoidable unfreedom. The market, moreover, is also an unsuitable mechanism for the distribution of *un*avoidable toil, since such toil, as a form of socially necessary unfreedom, constitutes a distinctive social burden incommensurable with socioeconomic goods.

For a number of reasons, then, I conclude that if we are genuinely committed to living in a society of free people and are entitled to hold one another to account for this commitment, we have a collective duty, and can make corresponding second-personal claims on another, to seek an alternative to the market as the central form of the social organization of productive activity. Whatever its advantages may be, the market is incompatible with the freedom of the individual.

4. THE RIGHT TO FREEDOM

Why give an account of freedom at all? As I have already hinted, the primary interest of an account of freedom for my purposes lies in its function as a social ideal: the ideal of a society of free people. It is this function which fixes my subject-matter.

I am comparatively uninterested in the role which the concept of freedom plays in our attributions of moral responsibility. This puts me at odds with much of the literature on freedom in moral and legal philosophy since at least Peter Strawson's "Freedom and Resentment."[10] I depart from the trend for two reasons. First, I suspect that the relevance of freedom for moral responsibility has been exaggerated. Generally speaking, at least claims of external duress, such as interpersonal coercion, function

---

[10] P. F. Strawson, "Freedom and Resentment," in *Freedom and Resentment and Other Essays* (London: Routledge, 2008).

primarily not to excuse action by denying or attenuating the agent's *responsibility* for it, but rather to justify action and thereby to negate or attenuate the agent's *culpability*.[11] It is true that a valid claim of duress can mitigate or negate some of the normal moral or legal consequences of an action—it can invalidate a contract or promise, for example—but not always, and in any case this is not all there is to responsibility.

Second, and more importantly, when we speak in political philosophy of the ideal of a society of free people, what we have in mind is (I hope) not in the first instance a society in which people can successfully hold one another morally responsible for their actions. This may be an upshot of living in a society of free people, but it is not most fundamentally what makes such a society desirable. I therefore think that, for political purposes, the focus on moral responsibility does not get to the heart of the matter.[12]

Given my focus on articulating the idea of freedom as social ideal, I presuppose throughout this dissertation that we are committed to living in a society of free people, and that we are entitled to hold one another to account for this commitment. I take it that a society of free people is one which secures for each of its members the right to individual freedom. In the first instance, this means freedom against coercion by other people. But as I have indicated, I will argue that the requirements of freedom go further, and are likely to demand a significant transformation of the basic structure of our

---

[11] See T. M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, MA: Harvard University Press, 2008), 180–181.
[12] Robert Nozick and Alfred Mele make the same point about "free will" and "autonomy," respectively. Robert Nozick, *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), 291; Alfred R. Mele, *Autonomous Agents: From Self-Control to Autonomy* (New York: Oxford University Press, 1995), 139.

society, particularly of the social organization of labor.

Now, I may seem to be proceeding in the wrong order. How can one reasonably begin by positing a commitment to freedom as a social ideal, or a right to freedom, *before* having given an account of what freedom is? I do not think this is as strange as it sounds. Generally, we come to political philosophy with pre-existing commitments to certain values, such as freedom and justice, as well as some sense of what they involve. Our sense of these commitments, however, is likely to be to some extent obscure, repressed, and misunderstood. It is a task of systematic political philosophy to bring to light, articulate, and either vindicate or correct these commitments. In this process, it is quite normal to begin with a pre-existing normative commitment whose content we do not yet fully understand.

I take it that this is roughly the idea behind John Rawls's method of reflective equilibrium.[13] Interestingly, though, Rawls himself takes a very different approach when it comes to the concept of freedom, as opposed to the concept of justice. When it comes to freedom, Rawls's approach is to define the concept by stipulation—as it happens, in terms of a small set of liberties from interpersonal interference in specific domains, such as the formation and expression of religious and moral beliefs, the choice of occupation, etc.—and then to argue that the citizens of a just society are entitled to an equal scheme of such liberties.[14] This is a common method in liberal political theory. And there is a

---

[13] John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 17–19, 40–46; *Justice as Fairness: A Restatement* (Cambridge, MA: Harvard University Press, 2001), 29–32. I am heavily indebted here to Ben Laurence's interpretation and defense of a "teleological" conception of political philosophy in Ben Laurence, *Agents of Change: Political Philosophy as Practical Reasoning* (Manuscript), especially Chapter 2.
[14] Rawls, *A Theory of Justice*, Chapter IV.

way in which it may seem to make more sense, since it seems not to rely on the independent presupposition of a pre-existing but imperfectly articulate commitment to the value and priority of freedom.[15]

But what this approach gains in the critical interrogation of the right to freedom, it purchases at the cost of an uncritical, stipulative method when it comes to the definition of freedom. This stipulative method is liable to lead us astray. First, the method of stipulation ignores the fact that freedom is not a technical term but a concept of ordinary thought and discourse, and a concept which shapes our pre-theoretical desires and political commitments. We cannot pretend that the word does not already have a meaning by the time we start philosophizing about it, or that it does not already play a role in structuring our desires and ideals. Indeed, even philosophers who employ the stipulative approach to defining freedom in effect borrow the aura of attractiveness which attaches to the word 'freedom' for the purpose of their technical definition. But if these technical definitions fail to capture what we really care about when it comes to freedom—what freedom really is—then they are not really entitled to this aura.

Second, if we define freedom by stipulation, we can easily end up asking the wrong questions. In a critical social theory, it matters what empirical concepts we use. We *can* define terms arbitrarily by stipulation—but if we begin with certain kinds of definition, we may end up failing to carve the social world at its joints, so to speak, and we may fail systematically to bring politically salient features of that world into view.

---

[15] Actually, this appearance may be misleading. At least in his earlier work, Rawls thinks of his theory of justice as a whole as articulating a Kantian ideal of autonomy and a conception of persons as free (and equal). Ibid., §40.

The same holds for normative concepts. We *can* define freedom arbitrarily by stipulation—for instance, as a set of specific liberties—and then consider whether we have or lack a right to freedom in this technical sense. But even if we successfully answer this specific question, our arbitrary definition of freedom may fail to capture an important value which the real concept of freedom captures; it may therefore leave important concerns unaddressed. Of course, this is only a possibility. Perhaps the definition will turn out to be fine. But this will have to be shown by argument, not by stipulation. My point, then, is that we cannot simply put aside or defer the question of what freedom really is in the hope that everything will turn out alright. We must tackle this question head-on.

CHAPTER I. IS FREEDOM ONE OR MANY?

There is no doubt that some "kinds" of freedom—particularly freedom from coercion by other people—are of special political significance. But are these kinds of freedom independently intelligible, or must we understand them as species of a single underlying genus—freedom, not in this or that respect, but *simpliciter*? In this methodological chapter, I make a start of answering this question by reflecting on what we should hope for and what we can expect from a philosophical account of freedom.

After describing some of the different species of freedom (Section 1), I argue that we should aim to explain in a non-trivial way both how these species differ from one another and what makes them all nevertheless species of a single genus (Section 2). I defend the possibility both of the generic unity of freedom and of a systematic philosophical account of this unity against general skeptical doubts (Sections 3–4). I then argue that there is no reason to assume in advance that "social" or "political" freedom corresponds to just one species of freedom (Section 5), before closing the chapter by indicating some limitations of my project (Section 6).

1. SPECIES OF FREEDOM AND UNFREEDOM

Freedom is said in many ways. We distinguish, among other things, between doing something freely and being free to do it, and between different sources of unfreedom.

1.1

We can speak of freedom under the aspect of actuality and under the aspect of potentiality. That is, we speak both of actually doing something freely and of being free to do something: of free activity and of free agency. As illustrated in Table 1, this basic grammatical distinction corresponds to a contrast in the categories of *un*freedom.

Table 1. Freedom as actuality and as potentiality

	Freedom	Unfreedom
Actuality	*ϕ-ing freely*	*ϕ-ing under force or compulsion*
Capacity	*Being free to ϕ*	*Being prevented or constrained from ϕ-ing*

To do something unfreely is to do it in some sense under force or compulsion: it is to be forced or compelled to do it. This contrast is associated with a contrast between desire and necessity. We speak of acting freely as a matter of doing what one *wants* to do, and of acting unfreely as a matter of doing what one *has* to do, where the relevant senses of these terms derive their meaning in part from their opposition to each other. On the other hand, to be unfree to do or undergo something is to be in some sense prevented or constrained from doing or undergoing it. And this contrast is associated with a contrast between ability or possibility on the one hand and inability or impossibility on the other.

Of course, to say all of this is not yet to explain what exactly words like 'force', 'compel', 'prevent', and 'constrain' mean in this context, or what the relevant concepts of desire, necessity, possibility, and impossibility come to here. That explanation is part

of what an account of freedom should give us. Presumably, such an account is also bound to show that the two registers of freedom and unfreedom are closely related. If I am forced or compelled into one course of action, it is generally to be expected that I am by the same token prevented or constrained from certain other courses of action.

1.2

In the ordinary run of things and in the absence of explicit elaboration or implicit contextual clarification, we do not simply ask whether someone is doing something freely (or free to do something), or assert that she is doing it freely (or free to do it), *simpliciter*. Rather, the question of freedom will ordinarily revolve around one or more of a variety of specific ways of being free or unfree—different kinds of freedom, so to speak. I call them the dimensions of freedom.

One way in which the concept of freedom ordinarily demands specification is in respect of the form of constraint or compulsion which is at issue—that is, the specific form of *un*freedom that is at issue. As Joel Feinberg observes,

> the word 'free', without further specification, is often incompletely informative. […] If a stranger casually informs you that he is free, you may have little idea of what he intends to convey until he tells you what he is free *from*. "Has he just escaped from prison, from his debts, from his wife, from his sins?" Until this further specification is forthcoming, you can infer only that he is asserting (with pleasure) the absence of *something* which he regards as an impediment or constraint.[1]

Normally, this will be obvious given the context. Absent either explicit or contextual clarification, however, it is generally strange simply to ask whether a given agent is, say,

---

[1] Joel Feinberg, *Social Philosophy* (Englewood Cliffs, N.J.: Prentice-Hall, 1973), 5, quoting Maurice Cranston, *Freedom, A New Analysis* (London: Longman, Green & Co., 1953).

doing something freely, or to announce that she is indeed doing something freely. Faced with such a general question or announcement, one will want to know what its more specific point is, that is, what specific forms of unfreedom are being ruled out.

For instance, suppose I tell you that a mutual friend of ours has undergone kidney surgery, and I add, "But she did it freely." Assuming it is not already clear what I have in mind, you will naturally expect to hear more at this point. What way of undergoing surgery unfreely am I intending to rule out? Was the surgery voluntary as opposed to coerced? (Perhaps our friend is in debt to some very bad people.) Was it a voluntary kidney donation as opposed to the medically necessary removal of a diseased organ? Was it a voluntary donation as opposed to the paid sale of an organ done under the force of economic necessity?

As a provisional heuristic, we can classify the various specific forms of compulsion and constraint to which freedom can be opposed in terms of two distinctions, which intersect to generate four possible forms of unfreedom, as illustrated in Table 2. Let us consider some examples for the purpose of illustration—on the understanding that this is bound to involve somewhat artificial simplification, since real-life unfreedom often does not fall neatly into just one of these categories.

Table 2. Dimensions of freedom

	Interpersonal	Non-interpersonal
External	*Physical coercion, volitional coercion*	*Natural disasters, public health crises, accidental pregnancies*
Internal	*Deception, brainwashing, torture*	*Phobias, addictions*

First, a given compulsion or constraint may be either interpersonal or not, that is (at a first pass), either imposed on one agent by another or not. The paradigmatic form of interpersonal compulsion is coercion by means of force or threats. In extreme cases, one agent can coerce another by directly controlling her body—for instance, by using her as a human shield against deadly projectiles. In this case, the coerced "action" is an action of the coerced agent in name only; as far as she is concerned, it is no more than a non-intentional movement. An agent can also coerce another intentionally do something, though, by threatening her with dire consequences unless she does it. This has been called "volitional" as opposed to "physical" coercion, since it is mediated by the coerced agent's own will. Slave labor extracted by means of the threat of physical violence would be a paradigm case, since such labor is performed intentionally but only because the worker risks a beating or worse if he refuses.

Equally, however, we sometimes speak of unfreedom when the source of compulsion or constraint is not interpersonal. Key examples are necessitation by one or another kind of accident for which no human agency is responsible. To cite a traditional example, a sailor who throws goods overboard in a storm to save herself and her ship can be said to do so unfreely, even if the storm happens to be an entirely non-anthropogenic natural phenomenon. Or again, if a pandemic forces people to stay at home, avoid socializing, and wash themselves and clean their environments incessantly, there is a sense in which they do these things unfreely. Or again, we might think that it makes sense to speak of the development of contraceptive technology as potentially emancipatory, because it brought a bodily condition which was previously forced on people to some extent by accident under their control at least in principle. Of course, the

actualization of this emancipatory potential depends on certain social conditions.[2] Indeed, more generally, unfreedom of a "non-interpersonal" kind is usually mediated in some way by social practices and institutions, so it is often misleading to describe it as merely "natural" as opposed to "social." Nevertheless, it is worth distinguishing from direct interpersonal coercion.

A second way in which we can distinguish possible forms of constraint and compulsion is based on whether they are "internal" or "external," or to put it another way, how directly they operate on the agent's will. External forms of compulsion, such as coercion by other people and compulsion by accidental circumstances, operate somewhat indirectly, by determining agents' environments or bodies in salient ways. For instance, when one agent coerces another to do something by threatening her, he induces her to do the thing by publicly changing the consequences of not doing it, *given* her existing ends or desires. Likewise, when a storm forces sailors to throw goods overboard, what happens is that the storm makes it the case that a sailor cannot pursue her existing ends—such as keeping the ship intact and surviving—without throwing goods overboard.

Other forms of compulsion, by contrast, are internal in the sense that they operate by directly shaping an agent's beliefs or desires themselves. This process, too, can be interpersonal. One example is deception. An agent can manipulate another into doing something by deceiving her about some pertinent matter of fact, for instance, by

---

[2] See Simone de Beauvoir, *The Second Sex*, trans. Constance Borde and Sheila Malovany-Chevallier (New York: Vintage, 2011), Vol. I, Part I (Chapter 1) and Part II, and Vol. II, Conclusion.

convincing her wrongly that she stands to gain from doing the thing. If she does the thing on account of having been deceived in this way, she acts unfreely. Manipulation can also target desires rather than beliefs. An agent can induce another to do something not only by threatening her or deceiving her about some matter of fact but by brainwashing her into wanting to do it. Clearly, it is intelligible to speak of an action as unfree insofar as it is the result of such brainwashing. Similarly, if a prisoner is forced to reveal a secret under torture, she will be acting under an internal form of compulsion—albeit one intentionally brought about by another agent—if she wishes she had the strength to resist but finds the pain so overwhelming that her desire to do whatever it takes to end it becomes irresistible.

On the other hand, compulsion through the direct distortion of an agent's will may also be non-interpersonal. Some phobias are a possible case in point: a person confined to her home by severe agoraphobia may reasonably feel that her condition makes her unfree. Some addictions may be like this also—although, as I have said, most real-world examples can be expected to evince multiple forms of unfreedom. Addiction, in particular, may be a response to objectively awful and seemingly inescapable circumstances. More generally, I would venture a guess that internal unfreedom tends to be parasitic on one or another kind of external unfreedom.

2. IN DEFENSE OF SYSTEMATICITY

In this section, I argue that a complete philosophical account of freedom should be systematic, in the sense of genuinely explaining both the unity of freedom and the multiplicity within this unity. On the one hand, it should be an account not merely of

freedom in this or that dimension but of the generic concept of freedom, or freedom *simpliciter*. It should characterize this genus in terms of a single principle capable of grounding all of the specific ways of being free or unfree. On the other hand, the account should explain not merely what the various species of freedom and unfreedom have in common but also what sets them apart from one another, in a way that goes beyond the superficial heuristic employed above. In particular, we will want an explanation of what is special, both morally and conceptually, about the important category of interpersonal coercion.

2.1

One aim of a systematic account of freedom is to identify and characterize what the various species of freedom have in common by virtue of partaking in a single common genus. The object of such an account, accordingly, is not in the first instance freedom in this or that specific dimension but freedom *simpliciter*. In addition to having this general object, moreover, a systematic account must explain this object in a genuinely general way. That is, it must characterize the genus of freedom in terms of a single principle rather than a series of fundamentally independent principles.

For the sake of illustration, what would it look like *not* to do this? The most important alternative would be to characterize freedom in terms of an irreducible (and perhaps open-ended) list of various ways of being free or various ways of not being unfree. This approach is associated with "compatibilism" in the literature on free will. Consider the following example of a non-systematic definition of freedom:

> An agent *S* [is free to] do *X* just in case (i) if *S* were to choose to do *X*, *S* would do *X*, and (ii) the agent is not subject to clandestine hypnosis, subliminal advertising, psychological compulsion resulting from past

> traumatic experiences, direct stimulation of the brain, neurological damage due to a fall or accident, and so forth…[3]

Presumably it is easy enough to show that each of the relevant elements of freedom—the absence of clandestine hypnosis, the absence of subliminal advertising, etc.—is consistent with the truth of "causal determinism." Thus, as Austin puts it, "[i]n examining all the ways in which an action may not be 'free', […] we may hope to dispose of the problem of freedom."[4] As the inevitable 'and so forth' highlights, such accounts define freedom essentially as a list of specific ways of being free (or not being unfree). A systematic account, by contrast, would identify more generally what all of these have in common.

Systematicity is not merely an aesthetic virtue but an essential condition of a critical understanding of ourselves and our world. The helpfulness of a philosophical account of freedom can be gauged by its explanatory power on the one hand and its critical and constructive power on the other. Though of course not all kinds of explanation are the proper aim of philosophy, an illuminating account of freedom still ought to deepen our understanding of our judgments of freedom and unfreedom in

---

[3] John Martin Fischer, "Compatibilism," in *Four Views on Free Will*, ed. Robert Kane, et al. (Malden: Blackwell, 2007), 51. The original quote reads "An agent *S* can do *X* just in case…," but Fischer uses 'can' synonymously with 'is free to'. Fischer himself, by the way, does not endorse this definition; he only mentions it in order to criticize it. A less extreme example of a non-systematic account would be Feinberg's initial characterization of autonomy in terms of an open-ended list of "remarkably miscellaneous" virtues (of which he names a dozen), "united only by a family resemblance, and a connection, however far removed, to the generating idea of self-government." Joel Feinberg, *The Moral Limits of the Criminal Law, Volume 3: Harm to Self* (New York: Oxford University Press, 1986), Chapter 18, Section 13.

[4] J. L. Austin, "A Plea for Excuses," in *Philosophical Papers*, ed. J. O. Urmson and G. J. Warnock (Oxford: Oxford University press, 1979), 180.

some non-trivial way. Moreover, it ought to give us some resources for interrogating those judgments rather than simply accepting them all uncritically, and for dealing with key controversial cases as well as extending our judgments to non-obvious cases. The generality of a systematic account, I will argue, is indispensable for all of these purposes, as it allows us to answer the crucial question *why* a given species of freedom or unfreedom is a species of freedom or unfreedom, if indeed it is.

Consider the explanatory dimension first. It is true that in severally examining the various species of freedom, whether in total isolation or as irreducible elements of a non-systematic account of generic freedom, we stand to deepen our understanding of each of these species individually. Moreover, we may learn much about the differences between them. This is not nothing—but it is not enough. For we wish to understand not just the differences between the many forms of freedom and unfreedom but also their unity, not just the differentiae of the various species but also their underlying genus. It is an understanding of this generic unity which allows us to answer the question why each of the various species is genuinely a form of freedom or unfreedom at all and not a mere homonym.

Clearly, a nonsystematic account cannot help us with this second inquiry. For according to such an account, the only basis for the determination whether a given case is a case of freedom or unfreedom is its presence or absence on the list of possible species. But it would contribute nothing to our understanding to say that a given case is a case of unfreedom because it is on this list, or not a case of unfreedom because it is not on the list. Only an account that appeals to something more general than the specific

kind of case under consideration can deepen our understanding on this front.[5] Because a systematic account of freedom defines freedom in terms of a general principle, it allows us to answer the question why the various cases of freedom are in fact cases of freedom—namely because they instantiate this general principle—and why the various cases of unfreedom are in fact cases of unfreedom—namely because they violate the principle.

Because of their limited ability to explain our judgments, non-systematic accounts also lack resources for criticizing them and for extending them to non-obvious cases. By contrast, it is here that a systematic account really comes into its own. As I have said, such an account can justify our existing judgments by showing how they follow from a more general principle—but by the same token it can also criticize them where they contradict that principle. For instance, should someone be inclined to treat one of the dimensions of freedom in Table 2 as a merely metaphorical, marginal, or downright confused use of the concept, we need not meet this inclination with either brute acquiescence on the one hand or a brute opposite inclination on the other. Instead, we can subject this inclination to critical scrutiny and either accept or reject it on general grounds. Neither the material in Table 2 nor any countervailing proposals amount to the

---

[5] Nozick convincingly argues that not all conjunctions of facts can be expected to have unified explanations as opposed to conjunctions of explanations. Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), 220–221. Fair enough: if two given facts are obviously completely unrelated to each other, there is no reason to demand a unified explanation of their conjunction. The situation is quite different, however, when the facts in question are obviously related—for instance in being instantiations or negations of one and the same concept. For instance, one does feel the need for a principle of unity to ground Nozick's second principle of justice, which regulates what he calls transfers of holdings: why are gifts, sales, wages, interest, etc. all permissible forms of transfer whereas theft, fraud, etc. are not?

last word on the matter.

As we have seen, non-systematic accounts of freedom can offer no serious answer to the question why any particular kind of case is or is not a case of unfreedom; their only guides for settling this question, should it arise, are the specific judgments we are inclined to make about the cases in question—or as they are often called in this context, "intuitions." But this means that the account has extremely limited resources for criticizing these judgments. One might try to elicit different judgments by pointing out misleading features of the cases under consideration or drawing attention to different comparisons.[6] One can even conduct surveys to show that many people share or dispute the judgments in question. What one cannot do, however, given a non-systematic account of freedom, is produce (good) reasons why we should judge cases one way or the other, and in particular why some of our specific judgments might be mistaken. In this sense, we must treat the judgments we are inclined to make about specific cases as finally authoritative.

For the same reasons, only a systematic account can help us extend our judgments to non-obvious cases. If we come across a case which we are not yet clearly inclined to judge one way or the other, a general principle gives us a way to come to a non-arbitrary conclusion about it. Since a non-systematic account, on the other hand, relies exclusively on our judgments about specific cases, it gives no guidance at all in those situations where we simply feel at a loss.

---

[6] Witness, for instance, Austin's delightful and instructive footnote about the difference between doing something "by accident" and doing it "by mistake." Austin, "A Plea for Excuses," 185n1.

The idea of systematicity in philosophy might seem quaintly "foundationalist." As Rawls argues, a good theory is answerable to considered judgments at all levels of generality, and none of these levels is uniquely privileged. In particular, the general is not always take precedence over the particular. If our theory of freedom implies, for instance, that slave labor is free activity, we know that we have gone wrong somewhere along the line and must revise some of the more general principles which led us to this conclusion. In the search for reflective equilibrium, as Rawls says, "we work from both ends."[7]

Properly understood, however, Rawlsian non-foundationalism is not merely consistent with but presupposes systematicity. For the answerability of general theoretical principles to specific considered judgments consists not merely in their consistency with those judgments but in their ability to make sense of those judgments, that is, to explain and justify them. We do not wish merely to be *able* to say that the slave acts unfreely; we wish to have good *reason* for saying this.[8] In the absence of such a reason, what we have is not a full-fledged judgment with the authority of knowledge so much as a brute inclination. A systematic understanding of freedom can be understood precisely as making such a reason explicit and thereby to endow what would otherwise be mere intuitions with the authority of knowledge.

---

[7] Rawls, *A Theory of Justice*, 18. See, more generally, ibid., 17–19, 40–46.; and *Justice as Fairness*, 29–32.

[8] Rawls himself is explicit about this at times. For instance, he thinks that the principles of a theory of justice "can serve as part of the premises of an argument which arrives at the matching [specific considered] judgments. We do not understand our sense of justice until we know in some systematic way covering a wide range of cases what these principles are." Rawls, *A Theory of Justice*, 41.

2.2

The other part of systematicity is difference. We want an account not just of the genus of freedom but also of the differentia of each of its species. In particular, there is something special about interpersonal coercion as opposed to other forms of unfreedom. As Rousseau observes, being confined to one's house by a snowstorm is one thing, and being deliberately locked in by another person is quite another.[9] In a distinctive way, it is "ill will" and not "the nature of things" which "maddens" us.[10]

To begin with, the interpersonal dimension of freedom or unfreedom has a distinctive moral and psychological significance. We tend to respond to coercion with a distinctive kind of interpersonal attitude—resentment—which remains absent in cases of "accidental" unfreedom.[11] This attitude is a response to a distinctive moral fact: the fact that, at least normally and defeasibly, someone who coerces us thereby *wrongs* us. The same moral fact gives rise in turn to an especially stringent kind of second-personal claim against others: against the coercer to stop coercing us, and against third parties to intervene to prevent or stop the coercer from coercing us. A fully systematic account of

---

[9] Jean-Jacques Rousseau, *Emile: or, On Education*, trans. Allan Bloom (Basic Books, 1979), cited, e.g., in David Miller, "Constraints on Freedom," *Ethics* 94 (1983): 69; and in Raymond Geuss, *History and Illusion in Politics* (Cambridge: Cambridge University Press, 2001), 104–105.

[10] Also a remark frequently attributed to Rousseau in the literature: e.g., Isaiah Berlin, "Two Concepts of Liberty," in *Four Essays on Liberty* (Oxford: Oxford University Press, 1969), 123; Miller, "Constraints on Freedom," 69; and Bernard Williams, "From Freedom to Liberty: The Construction of a Political Value," *Philosophy and Public Affairs* 30 (2001): 11.

[11] Numerous philosophers have emphasized this. It is a commonplace of the free-will literature since at least Strawson, "Freedom and Resentment." But see also, e.g., Williams, "From Freedom to Liberty."

freedom should put us in a position to understand this special moral significance of coercion in particular, without having to rely on implausible moral or psychological premises.

Furthermore, coercion seems to be special not just morally speaking but conceptually, as a form of unfreedom. That is, it seems that there is some distinctive way in which coercion subverts or bypasses its victim's will, where this difference is not accounted for just by its special etiology (though the etiology is useful as a basis for initial classifications). It is not just the cause of the unfreedom but the result—the quality of the will or the relation of the will to its actions—which is different in the case of coercion. Given only that someone is unfree in this sort of way, one already knows that this person's unfreedom is the result of coercion. Admittedly, this idea is somewhat vague and enigmatic, for the moment. Without some way of making sense of it, though, it may turn out to be more difficult than we imagine to make sense of what is special about coercion from a moral point of view.

3. THE POSSIBILITY OF GENERIC UNITY

Given the multiplicity of specific ways in which we speak of freedom, the pursuit of a systematic account of the genus must contend with two kinds of skepticism about this generic unity: skepticism about the existence of such a unity, and skepticism about its articulability. On the one hand, one might think that there simply is no single concept of freedom *simpliciter* but only various concepts of freedom of this or that specific kind. On the other hand, one might think that though there is such a thing as freedom *simpliciter*, it is impossible to give an account of this concept which is genuinely general as opposed

to something like an open-ended list of its species. In this section and the next, I flesh out the schematic ideal of systematicity by considering what it might look like for an account of freedom to answer both skeptical challenges. Drawing on the work of J. L. Austin—notwithstanding his own skeptical worries—I suggest that we conceive of the systematic unity of freedom on the model of "concepts of perfection" such as the concept of health.

3.1

The most important ground of skepticism about the generic unity, I think, is the supposition of a fundamental value pluralism at the heart of our aspirations for freedom.

When we speak of freedom in the course of everyday life, we are typically interested not merely in freedom *simpliciter* but in this or that specific way of being free or unfree. For our purposes in particular contexts, being free along just one such dimension may be sufficient to satisfy us that a given action counts as free. More to the point, the different dimensions are not always reconcilable but may come into conflict. For instance, we might think that someone engaging in extremely reckless behavior under the influence of a drug does not act freely. But it may be that the only way to stop her would involve some use of force, such as taking away her car keys, and of course we might think that this too subjects her to unfreedom of some kind. Safeguarding her freedom in one dimension, then, conflicts with safeguarding it in another. Absent some further specification as to which concern is paramount in the situation, one may feel that the question of how to safeguard her freedom *simpliciter* has no meaning in this context, and one may be tempted to conclude that there simply is no such thing.

Something like this reasoning appears to underlie many philosophers' conviction that, as Kristján Kristjánsson puts it, "there can be no general theory of *freedoms*. […] The most we may hope for is to clarify the concept of a specific *kind* of freedom."[12] Hayek makes it fairly explicit:

> the suggestion must be avoided that, because we employ the same word, these 'liberties' are different species of the same genus. […] [Other kinds of liberty] are not states of the same kind as individual liberty [from coercion]: we cannot, by sacrificing a little of the one in order to get more of the other, on balance gain some common element of freedom. […] If we have to choose between them, we cannot do so by asking whether liberty will be increased as a whole, but only by deciding which of these different states we value more highly.[13]

The existence of a single genus is denied, in other words, on the grounds that its supposed "species" can come into conflict with one another and that such conflict cannot be adjudicated by reference to a single generic metric.

Most influentially, this sort of thinking seems to motivate Isaiah Berlin's insistence that "positive liberty," or freedom as "self-mastery," is a fundamentally distinct concept from freedom as "negative liberty," or freedom as "non-interference" by others. On the face of it, this proposal seems preposterous—as though one could be one's own master while being subject to another's will, and as though the desire not to be subject to the will of others had nothing to do with the desire to be one's own master.[14] And indeed the proposal has attracted much criticism, perhaps most notably in

---

[12] Kristján Kristjánsson, *Social Freedom: The Responsibility View* (Cambridge: Cambridge University Press, 1996), 206. See also his treatment of the distinction between "social freedom" and "freedom of the will" on pp. 5–6.
[13] Friedrich A. Hayek, *The Constitution of Liberty* (Chicago: University of Chicago Press, 1960), 18–19.
[14] Berlin, "Two Concepts of Liberty."

the form of Gerald MacCallum's demonstration that Berlin's two "concepts" of liberty can be expressed in terms of a single schema.[15] But the reason Berlin and his followers have been unmoved by such criticisms, I suspect, is that the mere possibility of a common mode of expression does not undermine the allegation of a fundamental distinction in value. For political philosophers of a certain temperament, the aspiration for a unified theory of freedom rests on what Isaiah Berlin calls the "conviction that all the positive values in which men have believed must, in the end, be compatible, and perhaps even entail one another"—whereas in reality of course "not all good things are compatible, still less all the ideals of mankind."[16]

3.2

The possibility of tension between different species of freedom, however, by no means entails that they do not belong to a single genus or, in ethical terms, that they are not parts of a single, coherent ideal. For there is at least one promising way of making sense of this phenomenon in terms of a generic unity underlying the specific diversity which so impresses Berlin *et al*.

Consider a concept which does exhibit the requisite combination of unity and multiplicity: the concept of health. When we speak of someone as healthy in the course of everyday life, we normally have in mind the absence of some specific form or forms of unhealth. In the absence of explicit or contextual clues, it would generally be strange

---

[15] Gerald C. MacCallum, Jr., "Negative and Positive Freedom," *The Philosophical Review* 76, no. 3 (1967). For a perspicuous discussion, see Adam Swift, *Political Philosophy: A Beginners' Guide for Students and Politicians* (Cambridge: Polity Press, 2006), Part 2.
[16] Berlin, "Two Concepts of Liberty," 167. See also ibid., 131–132.

and disorienting for someone simply to describe themselves or another person as "healthy" or "unhealthy." Coming out of the blue, such a report would likely be met with a question like, "What do you mean?"

On different occasions, we may have any number of more specific things in mind. If someone has just recovered from an illness, we may communicate this fact by saying that she is healthy, without thereby suggesting that she is now free of all other ailments. Or again, we may certify someone's general fitness to be admitted to a physically demanding occupation such as being a battlefield soldier by saying that she is healthy, without thereby suggesting that she is currently free of passing illnesses like the flu. More generally, there are of course different areas of health, nowadays each under the charge of its own type of doctor, and we may take a distinctive interest in any of these on this or that occasion. Any medical specialist of one of these areas has the authority in principle to pronounce a patient "healthy," it being understood in the context that the claim is limited to this or that organ or bodily function.

Notwithstanding this multiplicity and specificity of the ordinary uses of the concept of health, however, there would appear to be a single concept underlying these uses. To be healthy *simpliciter*—or healthy as such, or perfectly healthy, or healthy without qualification—is to be free not just of one or another specific medical ailment but of all of them. It is not just to have a healthy heart, or a healthy stomach, or healthy eyes, but to be healthy in respect of all of one's organs and functions. No doubt any more positive definition of health will be more controversial and is bound to be somewhat schematic, but this does not mean that no such definition would be defensible. At a first pass, we could try saying that to be healthy *simpliciter* is to be in

such a condition, physically and mentally, as to be able to pursue a full range of human goods. We can understand this idea even if there never has been and never will be a completely healthy human being in this generic sense. The concept of health is a concept of perfection—an ideal point at the intersection of multiple axes in a single normative space, each axis corresponding to a specific way of living up to or falling away from the ideal.

Nor does the multiplicity and potential irreconcilability of normative *interests* which we can be expected to have in the different species of health undermine this underlying generic unity. In certain situations and for certain purposes, we may be more interested in some species of health than others. In certain situations, moreover, different species of health can turn out to be irreconcilable, such as when one has to subject oneself to dangerous and harmful chemotherapy in order to rid oneself of cancer. The problem here, however, lies not in the nature of the relevant value or values but in the unfortunate contingent situation. In some cases, it may not even be obvious which specific dimension of health is paramount or how to answer the question what best promotes our overall health—but obviously we would not infer that there is simply no such thing as our overall health, understood as an ideal condition involving the absence of both forms of unhealth. The problem is merely that it is unattainable in the circumstances. The idea of the genus remains intact, both purely conceptually and as a normative ideal, as long as the conflicting specific interests or values are not inherently contradictory but, on the contrary, parts or instantiations of a single more general interest or value. The definition of this underlying value may well be highly general, multiply realizable, and contested, and no doubt my proposed definition is inadequate,

but be that as it may, skepticism about the very idea of health *simpliciter*, in advance of rigorous inquiry, would be premature.

For all that has been shown so far, the concept of freedom, too, may be a concept of perfection. In practice, we typically restrict ourselves to speaking of being free in this or that specific way, from this or that specific constraint or compulsion, but this multiplicity is consistent with the possibility that there is such a thing as being free *simpliciter*, or being free as such, or being perfectly free, or being free without qualification—namely being free in all of these specific ways, and in particular, in both interpersonal and non-interpersonal ways at once. The specific interests we may have in the various dimensions of freedom, moreover, may be derivative of a single overarching interest we have in being free of all of them, and any conflicts between the various dimensions may be merely contingent, as they are in the example of the agent behaving recklessly under the influence of drugs.

As long as we can conceive of the concept of freedom as having this structure—an eminently reasonable hypothesis on the face of it—we cannot infer from the diversity of species and associated specific values that there is no such thing as a generic concept and value of freedom underlying these species. In advance of further argument, in other words, the first, fundamental kind of skepticism about the very object of a systematic account of freedom is unwarranted. Though there are important conceptual and normative differences between the species of freedom, it is possible to be *too* impressed by these differences.

4. THE POSSIBILITY OF SYSTEMATICITY

Even if one acknowledges the possibility of a single generic concept of freedom that underlies its various species, one may be skeptical of the possibility of articulating a general *account* of this concept, that is, of characterizing in a general way what it is that the various species of freedom have in common, what it is by virtue of which they are species of this concept. Examining what it would be for an account of freedom as a concept of perfection to answer this second form of skepticism will give us a more determinate idea of the possible shape of such an account.

4.1

J. L. Austin provides what amounts to a thorough exposition of the idea of freedom as a concept of perfection, but he is suspicious of the urge to seek an account of a single general "characteristic" which the various species of freedom have in common. In a passage in "A Plea for Excuses" which I have already partially quoted, he writes:

> Like 'real', 'free' is only used to rule out the suggestion of some or all of its recognized antitheses. As 'truth' is not a name for a characteristic of assertions, so 'freedom' is not a name for a characteristic of actions, but the name of a dimension in which actions are assessed. In examining all the ways in which an action may not be 'free', i.e. the cases in which it will not do to say simply 'X did A', we may hope to dispose of the problem of freedom.[17]

We can understand just what Austin means in light of his more extensive remarks on 'real' in *Sense and Sensibilia:*

> It is usually thought, and I dare say usually rightly thought, that what one might call the affirmative use of a term is basic—that, to understand 'x', we need to know what it is to be x, or to be an x, and that knowing this

[17] Austin, "A Plea for Excuses," 180.

> apprises us of what it is not to be x, not to be an x. But with 'real' […] it is the negative use that wears the trousers [*sic*]. That is, a definite sense attaches to the assertion that something is real, a real such-and-such, only in the light of a specific way in which it might be, or might have been, not real. 'A real duck' differs from the simple 'a duck' only in that it is used to exclude various ways of being not a real duck—but a dummy, a toy, a picture, a decoy, &c. and moreover I don't know just how to take the assertion that it's a real duck unless I know just what, on that particular occasion, the speaker has it in mind to exclude. This, of course, is why the attempt to find a characteristic common to all things that are or could be called 'real' is doomed to failure; the function of 'real' is not to contribute positively to the characterization of anything, but to exclude possible ways of being not real—and these ways are both numerous for particular kinds of things, and liable to be quite different for things of different kinds. It is this identity of general function combined with immense diversity of specific applications which gives to the word 'real' the, at first sight, baffling feature of having neither one single 'meaning', nor yet ambiguity, a number of different meanings.[18]

How does being a real duck differ from being a duck? Austin's thought is: in a sense, not at all. In calling a duck real, we do not attribute to it some characteristic not already implied by its being a duck. Rather, we rule out one or more specific ways in which it might be thought to be *not* (quite) a duck. Reality, on Austin's account, is a concept of perfection like the concept of health, and his discussion illuminates an important feature of such concepts.

Consider again the concept of health. A healthy duck, like a real duck, is not a duck with some additional characteristic, but simply a duck without qualification. One can describe a healthy duck just by describing a duck—and if one described what it is to be a duck, one would effectively be describing a healthy duck. To call a duck healthy, accordingly, is not to attribute some additional characteristic to it but only to rule out various ways in which it might be (or might be thought to be) unhealthy, that is, for lack

---

[18] *Sense and Sensibilia* (Oxford: Clarendon Press, 1962), 70–71.

of a better word, imperfect *qua* duck. A simpler word which can function in the same way is the word 'pure'.[19] To call gold pure is not to attribute to it some additional characteristic of "purity" but rather to rule out various forms of impurity. It is to say, in other words, that it is gold without qualification, gold in the strictest sense. Pure gold meets a standard of perfection of which impure gold falls short.

These observations suggest that a full account of a concept of perfection must do two things. First, it must display the conceptual structure just described: to apply such a concept to a thing is to say that the thing is, in some space or spaces of assessment, a perfect or unqualified instance of its kind. But this is not enough. For there are many possible such spaces, and a thing may be perfect with respect to some of them while falling short with respect to others. For instance, a duck may be real without being healthy—such as an injured duck in Central Park—or healthy without being real—such as Donald Duck in his prime. Second, therefore, a full account of one of these concepts must pick out the salient space or spaces of assessment that the concept brings into play.

It is arguably here that we find the real source of Austin's commitment to non-systematicity about such concepts. As he points out, there is an immense multiplicity of ways in which things may be, for instance, real or unreal. This multiplicity can be divided into two kinds, and Austin evidently thinks that either kind is sufficient to defeat any attempt to arrive at a non-trivially general characterization of what it is for a thing to be real or unreal.

---

[19] Anton Ford, "Action and Generality," in *Essays on Anscombe's Intention*, ed. Anton Ford, Jennifer Hornsby, and Frederick Stoutland (Cambridge, MA: Harvard University Press, 2011), 90–94.

First, things of different kinds may be real or unreal in many ways. For instance, cream may be real as opposed to artificial, but this opposition does not quite make sense for ducks; a duck, on the other hand, may be real as opposed to a decoy, but this opposition does not apply to cream; and the color of someone's hair may be real as opposed to dyed, an opposition which does not apply to either cream or ducks. But it may be no easy matter—indeed, it seems hopeless—to say what being artificial (cream), being a decoy (duck), being dyed (hair color), etc. have in common. It seems that if we are to give a philosophical account of 'real'—of "the Nature of Reality," as Austin says tongue-in-cheek—we must first specify the kind of thing whose reality or unreality is in question. On this view, even given an account of what it is to be a real duck, we could not formulate a systematic account of what it is to be real *simpliciter*. At best we could put together something like an open-ended list of ways of being real: *to be real is to be a real duck or real cream or ….*

But second, even things of a given kind may be real or unreal in many ways. A duck, for instance, may be real as opposed to a decoy, real as opposed to a toy, real as opposed to a hallucination, or real as opposed to fictional. What do being a decoy, being a toy, being a hallucination, and being fictional have in common? Despairing of an answer to this kind of question, Austin thinks that the best we can do is to examine severally all the ways in which a duck may not be real, i.e., the cases in which it will not do to say simply, "X is a duck." This is not to say that there is no such thing as being a real duck *simpliciter*, since it is not under dispute that there is such a thing as being a duck, period. It does imply, however, that a full philosophical account of what it is to be a real duck—an account which identified the dimension of assessment proper to this

particular concept of perfection—could once again only take the form of a list of fundamentally independent elements: *a duck is real unless it is a decoy or a hallucination or …*.

Now we can see the full point of Austin's analogy between 'real' and 'free'. In calling an action free, he thinks, we do not attribute some further characteristic to it but rather rule out various ways in which it might have failed to be free. A free action is not an action with some further attribute lacked by unfree actions, but simply an action without qualification; it meets a standard of perfection of which unfree actions fall short. Accordingly, a full account of freedom must both display this structure and identify the distinctive dimension or dimensions of perfection or imperfection brought into play by this concept. Austin does not deny that there is such a thing as freedom *simpliciter*; he admits, after all, that to call an action free may be to rule out "*all* of its antitheses" (emphasis mine). His skeptical claim is rather that we cannot identify the relevant dimension or dimensions of assessment except in a negative and disjunctive way, by listing the various ways in which an action may be *un*free: *an action is unfree if … or … or …*; or: *an action is free unless … or … or …*.

4.2

Is Austin's skepticism warranted? Must we think that there is no genuinely general characterization of the relevant dimension of perfection? As in the case of the concept of reality, there are two possible grounds for pessimism. First, one might reasonably despair of giving a single, systematic account of freedom that encompasses the concept in its application to all kinds of things, from *free action* through *free lunch* to *free radical*.

But this merely suggests that we should begin not with the idea of freedom in abstraction from any particular subject, but with something more determinate, such as the idea of free activity or free agency. The important question is whether we have a reasonable hope of giving an account of freedom in this particular connection.

At this level, however—that is, once we have specified the kind of thing to which the relevant concept of perfection is to be applied—skepticism about philosophical systematicity cannot be the default presumption. Austin may well happen to be right about the concept of reality. But consider, say, the concept of purity. In abstraction from any particular subject, the attempt to give a systematic account of this concept would seem fairly hopeless. Once a subject is specified, however—gold, for instance—the situation is altogether different. Though there are indefinitely many different kinds of contaminant which may render gold impure, and as many "ways" in which gold may be pure, it clearly does not follow that we cannot give a non-trivially general account of what it is for gold to be pure, one which both displays the structure of this concept and specifies the relevant dimension of assessment. Pure gold is gold which is not mixed or compounded with any other substance. Indeed, the generic purity of gold thus understood even admits of measurement, namely as the ratio of the mass of gold within a given object to the total mass of the object.

Absent quite general grounds for skepticism, it is unclear why we should presume that in the special case of freedom, no systematic account of the concept is possible even once we have specified that it is free activity or free agency that is in question. Arguably, we already have some clues for where to look for such an account. As a starting point, in roughly classifying various forms of unfreedom above, I identified all of them as so many forms of constraint or compulsion—ways in which

someone may be either forced to do something or prevented from doing it, and thus ways in which activity or agency may be subject to some kind of necessity or impossibility.

Now, not *all* forms of constraint or compulsion, force or prevention, and necessity or impossibility are reasonably regarded as forms of unfreedom. More will therefore need to be said to specify what makes a constraint or compulsion a source of unfreedom. Moreover, the absence of constraint or compulsion remains a purely negative characterization of freedom, and we still lack a more positive characterization of the concept. Finally, as I said, we will ultimately want to distinguish different forms of constraint and compulsion on the basis of more than their etiology alone. None of this, however, is grounds for assuming—in advance of making an actual attempt—that the enterprise is necessarily doomed. At this early stage, we ought instead to proceed on the presumption that a systematic account of free activity or agency is possible unless we actually encounter insurmountable difficulties. The proof of this presumption will be in the pudding.

5. Social relations and the depoliticization of nature

There is another way of exaggerating the differences between the dimensions of freedom, and the distinctiveness of interpersonal coercion vis-à-vis other forms of unfreedom in particular, namely by depoliticizing the other sorts of unfreedom. In this section, I want to caution against the premature assumption that it is only the interpersonal dimension of freedom which can be relevant from a specifically social or political point of view.

5.1

Philosophers frequently reserve the label of "social freedom" or "political freedom" for freedom in its interpersonal dimension. "You lack political liberty or freedom," in Berlin's words, "only if you are prevented from attaining a goal by human beings. Mere incapacity to attain a goal is not lack of political freedom."[20]

Now, sometimes this terminological convention is introduced as no more than a convenience or tautology, a way of referring to the distinctive source of this unfreedom in the agency of other people. Often, though, the label comes to be used to imply something more, not about the *source* of such unfreedom but about its *significance* or *implications*. It is implied (or explicitly said), specifically, that only freedom in its interpersonal dimension is of importance in the context of social or political theory.

"Ethics and metaphysics," as Joel Feinberg puts it, "may quite properly concern themselves with more varied sorts of constraints," but "political and social philosophy are concerned with freedom only when conceived as the absence of coercion by others."[21] One tends to find this sort of idea most explicitly in libertarian tracts, both right and left, such as in the work of Hayek, Robert Nozick, Milton and Rose Friedman, and Hillel Steiner.[22] To highlight one example, the Friedmans write:

> As liberals, we take freedom of the individual, or perhaps the family, as our ultimate goal in judging social arrangements. Freedom as a value in this sense has to do with the interrelations among people; it has no

---

[20] Berlin, "Two Concepts of Liberty," 122.
[21] Feinberg, *Social Philosophy*, 9.
[22] See Hayek, *The Constitution of Liberty*, 11, 16; Milton Friedman and Rose D. Friedman, *Capitalism and Freedom* (Chicago: University of Chicago Press, 1962), 12; Nozick, *Anarchy, State, and Utopia*, 262–264; Hillel Steiner, "Individual Liberty," *Proceedings of the Aristotelian Society* 75 (1974–1975).

> meaning whatsoever to a Robinson Crusoe on an isolated island (without his Man Friday). Robinson Crusoe on his island is subject to "constraint," he has limited "power," and he has only a limited number of alternatives, but there is no problem of freedom in the sense that is relevant to our discussion. […] There are thus two sets of values that a liberal will emphasize—the values that are relevant to relations among people, which is the context in which he assigns first priority to freedom; and the values that are relevant to the individual in the exercise of his freedom, which is the realm of individual ethics and philosophy.[23]

The idea also makes its way, though, into more purely "conceptual" investigations of freedom, such as in work by Felix Oppenheim and Kristjánsson.[24]

Although it is often put forward as though it were self-evident, the assertion that only interpersonal freedom has "social" or "political" significance is not a tautology but a substantive political claim.[25] What substantive grounds, then, might one give for this claim? Presumably it is based on some conception of the meaning of the distinction between the "social" and the "natural." In Pettit's words,

> the distinction between securing people against the natural effects of chance and incapacity and scarcity and securing them against the things that they may try to do to one another […] is of first importance in political philosophy, and almost all traditions have marked it by associating a person's freedom with constraints only on more or less intentional interventions by others […].[26]

What precisely is the significance of this distinction? I have already mentioned that it is

---

[23] Friedman and Friedman, *Capitalism and Freedom*, 12.
[24] E.g., Felix E. Oppenheim, *Dimensions of Freedom* (New York: St. Martin's Press, 1961), 109; and Kristjánsson, *Social Freedom*.
[25] We must certainly resist the all-too-easy slide from the tautological function of the terminology of "social freedom" to its other, substantive function. To illustrate: Kristjánsson, having defined "social freedom" in strictly interpersonal terms, asks rhetorically: "what sort of freedom is relevant to political question[s] if not *social* freedom[?]" *Social Freedom*, 20, emphasis his.
[26] Philip Pettit, *Republicanism: A Theory of Freedom and Government* (Oxford: Oxford University Press, 1997), 52–53.

associated with important attitudinal differences: we *resent* interpersonal unfreedom in a way in which we do not resent facts of nature. More important, though, are the distinctive *practical* implications of interpersonal unfreedom. Specifically, it gives rise to immediate and especially stringent second-personal claims against others. Indeed, it is easy to suppose, more strongly, that non-interpersonal unfreedom cannot give rise to second-personal claims at all, and this can in turn seem to make it socially or politically insignificant.

Consider what Elizabeth Anderson says, following Stephen Darwall, about the concept of justice:

> a judgment of justice is essentially expressible as a *claim* or demand, made by or on behalf of someone, on another agent's conduct, directing that agent to serve, or not injure or neglect, the interests of the person on whose behalf the claim is made. A judgment of injustice is essentially expressible as a *complaint* addressed to an agent, who is held accountable to the person making the complaint, about that agent's failure to comply with valid demands that the agent serve or pay due regard to the interests of the claimant.[27]

We might think that in the context of social and political philosophy, judgments concerning freedom and unfreedom must likewise be capable of grounding not merely judgments of goodness but second-personal claims and complaints. That is, a judgment of freedom must be capable of grounding not merely a judgment to the effect that a certain kind of capability, or liberation from a certain kind of constraint or compulsion, would be good; it must be capable of grounding a demand on another agent's conduct,

---

[27] Elizabeth Anderson, "The Fundamental Disagreement Between Luck Egalitarians and Relational Egalitarians," *Canadian Journal of Philosophy* Supp. Vol. 36 (2010): 4, emphasis hers. See Stephen Darwall, *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge, MA: Harvard University Press, 2006).

directing her to take steps to support the relevant capability or to remove the relevant constraint or compulsion. Likewise, a judgment of unfreedom must be capable of grounding a complaint to another agent, holding her accountable for imposing or failing to remove the relevant constraint or compulsion.

But it might seem that judgments of freedom and unfreedom could not have this distinctive social-practical force at all—could not be capable of grounding any second-personal claims or complaints—unless freedom itself constituted one or another kind of second-personal relation, and unfreedom a defective form of such relation. Interpersonal relations of coercion themselves constitute defects in our relations to one another, so it does not seem problematic that they should be capable of grounding second-personal complaints and corresponding claims on one another's conduct. Such claims would be grounded in the very interpersonal relations of which the unfreedom in question is a defect. By contrast, a merely "natural" constraint or compulsion may seem not to constitute a defect in our social relations, or indeed relate us to one another in any way at all. It therefore seems mysterious how it could license a second-personal complaint, or what the ground could be of a corresponding claim on anyone else's conduct directing them to eliminate the constraint or compulsion in question.[28] At best, the agent in question might have a claim or complaint against God.

5.2

But the opposition between the social world and the natural world which this argument

---

[28] Compare Miller, "Constraints on Freedom," 69; and Williams, "From Freedom to Liberty," 12.

presupposes is a false dichotomy.

Our relations to nature are not generally distinct from our social world but may be partly constitutive of it. This will be the case whenever a given set of social relations—a given community—is partly defined by shared activities that essentially relate its members jointly to nature in some determinate way. Suppose that we are members of a cooperative productive scheme. As members of a community defined by our shared productive activity, we relate ourselves to one another partly through our relations to nature. Our social positions and interpersonal relations, whether of hierarchy or of equality, as well as more generally what it means to be a member of this community at all, will be defined in part by reference to various parts of the extrapersonal world, such as the raw materials, workspaces, instruments, and products of our activity.

If we are members of such a community and our relations to nature break down in ways salient to our shared activity, our common membership of the community is sufficient to ground relevant second-personal claims. Suppose that there is a shortage of some particular raw material or that a natural disaster renders some of our workspaces or instruments unusable, and suppose that this calamity is not caused by anyone's actions or omissions. Even so, it concerns us *qua* participants of our joint cooperative scheme—that is, by virtue of what it is to be a member of the relevant community at all and not merely by virtue of some further value accidental to this membership.

Moreover, the event can conceivably put us in a position to make claims on one another's conduct, directing one another to take steps to address the relevant problem or to help us to do so. Though the ground of these second-personal claims lies within

and not beyond our social relations, the form of these grounding relations is expressed in this case not in the second person but in the first-person plural. *I* can make a claim on *you* to take relevant steps to address the "natural" calamity because *we* are engaged in a joint activity threatened by this calamity.

It is not at all clear what it would look like for a community not to involve such relations to the extrapersonal world. To relate to other people under the aegis of shared ends normally is to enter into such relations, too. This is normally true even when the end in question consists only in relating to one another in a certain way. For our strictly interpersonal relations, too, rely on the maintenance of certain relations to nature, and can be threatened when the latter break down. For instance, even if our shared end is just the interpersonal end of living together in relations of social equality, this will give rise to general second-personal claims to protect and promote whatever capabilities are necessary to sustain such interpersonal relations—and these may be threatened and supported not just by intentional human action but also by various extrapersonal material conditions.[29]

In general, then, the idea that social freedom is strictly interpersonal presupposes a blinkered vision of our social world, one which artificially abstracts from the relations to extrapersonal nature normally implicated in our practical relations to one another. Freedom is social indeed—but the natural world is generally part of the social world.

It is true that when our freedom is undermined not by another's action but by an "act of God," we have no grounds for a second-personal *complaint*. We may come to

---

[29] See Elizabeth Anderson, "What is the Point of Equality?," *Ethics* 109, no. 2 (1999): 316.

have such grounds if and when people on whom we now have claims fail to honor those claims, but these complaints would be posterior to the relevant claims, both conceptually and in time. We are not *wronged* by the initial "act of God" itself. For this reason, the claims we have against others to assist in addressing the relevant problem are likely to be less stringent and urgent than the claims we would have against them to cease and desist from coercing us, or to prevent third parties from coercing us. But it does not follow that we have no claims at all in this situation. There is no reason to assume that second-personal claims must always be based on prior second-personal complaints.[30]

I am rejecting, in effect, the common depiction of the ideal of freedom *simpliciter* as somehow essentially asocial, as an ideal essentially for people construed as isolated individuals on the model of Robinson Crusoe.[31] It is true that the ideal of freedom *simpliciter* admits of some kind of application to Robinson Crusoe, but it does not follow that this is its proper sphere of application. As a matter of fact, Crusoe is also free of interpersonal coercion on his island.[32] It obviously does not follow that his island is the

---

[30] This is why it is equally arbitrary to limit "social" freedom to freedom in relation to constraints and compulsions for which no other person is *morally* (as opposed to causally) responsible. See, e.g., S. I. Benn and W. L. Weinstein, "Being Free to Act and Being a Free Man," *Mind* 80, no. 2 (1971); Kristjánsson, *Social Freedom*; and Andreas T. Schmidt, "Abilities and the Sources of Unfreedom," *Ethics* 127 (2016).

[31] For another example of this sort of depiction, see Arthur Ripstein, *Force and Freedom: Kant's Legal and Political Philosophy* (Cambridge, MA: Harvard University Press, 2009), 15.

[32] If we take his story literally, that is. In the first sequel to the original book, Crusoe reveals that "all these Reflections are just History of a State of forc'd Confinement, which in my Real History is represented by a confin'd Retreat in an Island; and 'tis as reasonable to represent one kind of Imprisonment by another, as it is to represent any

proper sphere of application of the idea of interpersonal freedom. Of course, interpersonal theorists of freedom might reply that theirs is not merely a negative ideal of an absence but a positive ideal of social relations constituted in a certain way. But there is no good reason why this positive ideal should be limited to interpersonal relations only.

I sympathize with the interpersonal theorists' desire to leave Crusoe's wretched island. But when we move—or rather, return—from this island to society, we do not merely add interpersonal relations to our essentially asocial relations to nature. Rather, our relations to nature are transformed (back) into social relations. Acts of God now acquire a new (or rather, their full and proper) significance: in principle, they concern us not merely as individuals but as members of a community. My complaint about purely interpersonal theories of freedom, then, is not that they are too relational. They are not relational enough.

6. LIMITATIONS OF MY PROJECT

Having defended an ambitious ideal of systematicity, it is now time for me to lower expectations. I will not be defending a complete account of freedom in this dissertation. My subject-matter will be more limited in two ways.

First, I will focus almost exclusively on the external dimension of freedom, in both its interpersonal and non-interpersonal forms, with an emphasis on the interpersonal. My immediate topic, in other words, will be freedom in relation to other

---

Thing that really exists, by that which exists not." Daniel Defoe, *Robinson Crusoe* (Oxford: Oxford University Press, 2007), 267.

people and the extrapersonal material world. Tackling the fraught issue of internal freedom as well would simply be too large an undertaking. Of course, given that there are different species of external freedom, considerations of systematicity will still be relevant. Moreover, I will be working on the hypothesis that the account I give can be incorporated without substantial modifications into a more complete whole. I am thus not abandoning the ideal of systematicity, but merely defending one part of what I envision as a systematic account of freedom as a whole. In the absence of the remainder, this part accordingly remains somewhat provisional.

Second, because I am ultimately interested especially in the social organization of labor—an *activity* which can be done freely or unfreely—my primary topic will be freedom under the aspect of actuality. That is, my guiding question will be what it is to do something freely, rather than what it is to be free to do something. Of course, it is to be hoped that an answer to the one question would help us answer the other. Be that as it may, I will mostly talk about the former.

Freedom in this register raises a general problem, namely that one can make another person free to do something but one cannot in general make another person do it freely. After all, *making* her do it will normally be a way for her to do it unfreely. "One can bring the horse to the water," as Joseph Raz summarizes the point, "but one cannot make it drink."[33] This means that at some point—in favorable conditions—each person becomes responsible for their own freedom. What others can do is help create the social and material conditions of the possibility of free activity—even if the value of this

---

[33] Joseph Raz, *The Morality of Freedom* (Oxford: Oxford University Press, 1986), 407.

possibility is parasitic on the value of its actualization.[34]

Notwithstanding my focus on free activity, then, I will assume that our political duties will primarily concern promoting the *opportunity* for free activity. One way of doing this is of course to refrain from and to prevent coercion. In addition, philosophers from Mill to Raz have emphasized the duty to help create "the inner capacities required for the conduct of an autonomous life" and "an adequate range of options […] to choose from."[35] Going beyond the content of these traditional duties, I will end up arguing that securing the possibility of autonomy for all people also requires the social reorganization of labor.

---

[34] Ibid., 407–409.
[35] Ibid., 407–408.

CHAPTER II. KANTIANISM, MORALISM, AND PSYCHOLOGISM

In this chapter, I survey the main extant theories of freedom and unfreedom, focusing especially on how these theories try to make sense of coercion. Some of the theories are grounded in conceptions of freedom as autonomy; others attempt to theorize coercion entirely on its own terms. Some are relational, understanding freedom exclusively in terms of interpersonal relations; others understand it as also encompassing relations of persons to the extrapersonal world. For my purposes, I will divide the theories into three families: Kantianism, moralism, and psychologism. Throughout the chapter, I will often ignore differences within each of these families insofar as these are not relevant to my arguments.

According to Kantianism, freedom is autonomy, which is a matter of the unimpeded, unimpaired exercise of practical reason, and coercion is wrong because it subverts or bypasses this capacity in another (Section 1). I argue that at least in its dominant contemporary form, this view fails to do justice to the fundamental inalienability of agency—the fact that even the coerced agent exercises her practical reason and makes genuine choices in the pursuit of her ends (Section 2). Moralism, on the other hand, attempts to explain what is wrong with coercion not in terms of the subversion of the coercee's autonomy but in terms of the violation of some independent right of hers (Section 3). I argue that this account reverses the proper order of priority between freedom and the relevant kind of right (Section 4). Finally, according to psychologism, freedom consists in autonomy understood as the coherence of action

with desire and of first-order desires with higher-order desires, and coercion is wrongful, roughly speaking, because we generally have a higher-order desire not to be motivated in this sort of way (Section 5). I argue that this account reverses the proper order of priority between freedom and the relevant higher-order desires (Section 6). I suggest, though, that we might be able to salvage an important truth in psychologism by beginning with a different conception of first-order desire (Section 7).

1. KANTIANISM

Perhaps the most important modern tradition of thought about freedom is the Kantian tradition, which identifies freedom with self-determination, or autonomy.

Autonomy, for Kant, is the self-government of the will, understood as the power of practical reason. In the *Groundwork of the Metaphysics of Morals*, Kant opposes such self-determination principally to an internal kind of unfreedom which takes the form of the alien determination of the will by mere "inclinations," desires governed not by the will's own self-given laws but by the laws of nature.[1] But the Kantian idea of freedom as autonomy also naturally suggests an account of external unfreedom, at least in relation to other people. To coerce another person, on this account, is to use them as a mere means. This allows them only a privative exercise of their agency, in the service of alien ends rather than their own. And such treatment is wrong precisely because it undermines the other's freedom in this way. It refuses to recognize the other's status as

[1] Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1998), 4:446–447.

an agent in their own right, treating them instead as a mere tool for the use of others.

Let us see how this account of coercion might work. Kant's categorical imperative enjoins us "[s]o [to] act that you use your humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means."[2] As Kantians have interpreted it, "humanity" here means practical reason, or that aspect of it which consists in the ability to determine our own ends.[3] The idea is commonly spelled out in terms of the distinction between persons and things. As Christine Korsgaard puts it:

> A person, an end in itself, is a free cause, which is to say a first cause. By contrast, a thing, a means is a merely mediate cause, a link in the chain. A first cause is, obviously, the initiator of a causal chain, hence a real determiner of what will happen. The idea of deciding for yourself whether you will contribute to a given end can be represented as a decision whether to initiate that causal chain which constitutes your contribution.[4]

When we set out to get other people to do things we want them to do, we must do so in a way which respects their possession of this power:

> To treat others as ends in themselves is always to address and deal with them as rational beings. Every rational beings gets to reason out, for herself, what she is to think, choose, or do. So if you need someone's contribution to your end, you must put the facts before her and ask for her contribution.[5]

When one coerces someone, on the other hand, one secures her contribution without

---

[2] Ibid., 429, italics removed.
[3] See Christine M. Korsgaard, "Kant's Formula of Humanity," in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), esp. 110–114, for an exegetical argument.
[4] "The Right to Lie: Kant on Dealing with Evil," 140–141.
[5] Ibid., 142.

giving her a chance to either consent or dissent.[6] Coercion is thus a way of "taking a decision out of someone's hands"—a way of treating her "as a mediate rather than a first cause; hence as a mere means, a thing, a tool."[7] And this is wrong because it fails to respect her status as a practical reasoner. It is a way of undermining the other's autonomy.

Other Kantians use much the same language to make much the same point. Onora O'Neill says that coercion "denies [its victims] the choice between consent and dissent."[8] Barbara Herman defines autonomy as the capacity of agents to "fully determine their actions according to reasons," or as she also puts it, "to [act] for reasons 'all the way down'."[9] She takes this to explain why deception is wrong:

> The maxim of deception places the grounds of choice-worthiness of the victim's reasons in the deceiver's will. […] In effect, the deceiver has a maxim of treating a rational agent as someone whose will may be brought under causal control—as one whose reasons do not go all the way down. […] It violates the requirement that we regard reasons as coming to an end in the will of each agent, separately.[10]

Presumably, Herman's account of coercion would look similar. Arthur Ripstein, defending what he calls the Kantian idea of freedom as "independence," draws on the same distinction as Korsgaard:

> The core idea of independence is an articulation of the distinction between persons and things. A person is a being capable of setting his or her own purposes, while a thing is something that can be used in pursuit of

---

[6] "People cannot assent to a way of acting when they are given no chance to do so. The most obvious instance of this is when coercion is used." Ibid., 138, italics removed.
[7] Ibid., 142.
[8] Onora O'Neill, "Between Consenting Adults," *Philosophy and Public Affairs* 14, no. 3 (1985): 262.
[9] Barbara Herman, "Leaving Deontology Behind," in *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993), 228.
[10] Ibid., 229–230.

> purposes. […] You are independent if you are the one who decides which purposes you will pursue. […] You are dependent on another person's choice if that person gets to decide what purposes you will pursue.[11]

Later he concludes:

> The slave's problem is that he is subject to the master's choice: the master gets to decide what to do with the slave and what the slave will do. The slave does not set his own ends, but is merely a means for ends set by someone else.[12]

Though this approach is associated most strongly with Kantian moral philosophy (and I will continue to call it Kantian), it is not exclusive to that tradition. It is reminiscent, of course, of Aristotle's definition of the slave as "a living instrument" (though, lacking a modern universalist conception of rational or even human agency, he failed to see this as a reason to reject slavery).[13] More recently, the same approach crops up in some libertarian writings. Hayek, for instance, defines coercion as follows:

> By 'coercion' we mean such control of the environment or circumstances of a person by another that, in order to avoid greater evil, he is forced to act not according to a coherent plan of his own but to serve the ends of another. […] Coercion is evil precisely because it thus eliminates an individual as a thinking and valuing person and makes him a bare tool in the achievement of the ends of another.[14]

This is, in short, a very natural and compelling way of thinking about coercion, and one feels that it must be on the right track.

2. The fundamental inalienability of agency

In the absence of further arguments and distinctions, however, the Kantian approach

---

[11] Ripstein, *Force and Freedom*, 14–15.
[12] Ibid., 36.
[13] Aristotle, *Politics*, , 1243b–1254a.
[14] Hayek, *The Constitution of Liberty*, 20–21.

fails to explain both the unfreedom of coerced action and the wrongfulness of coercion. The account fails to do justice to the fundamental inalienability of agency—the fact that even unfree action is a genuine exercise of practical reason on the part of its agent.

2.1

There are really at least two kinds of coercion: physical coercion, which is direct coercion of the body, and volitional coercion, which is coercion of the will. An example of physical coercion would be one person grabbing another and wielding the other's body as a human shield against deadly projectiles. It is easy to see how the Kantian approach can help us understand this kind of coercion. If one agent overpowers another and moves their body into harm's way, the overpowered agent does not choose to move her body in this way; someone else takes the choice out of her hands and makes it for her. Her movement is thus not the result of her own practical reasoning, and it is an "action" in name only.

Most coercion, however, is not like this. Most coercion is mediated in some way by the coerced agent's own will. In general, volitional coercion works not by directly applying force to someone's body but by using a threat in order to get them to do what one wants them to do. The coercer makes it the case that if the coercee refuses to obey, she will suffer some undesired consequence, so that if she is to avoid that consequence, she will have to obey. But here we have a problem. Merely threatening someone in this way does not actually, literally deprive her of the ability to choose how to act. It may be very undesirable for her to refuse to obey, given the threatened consequence—but that does not make her literally unable to refuse. Refusal and obedience alike are still fundamentally her choice. In the case of volitional coercion, then, it seems simply false

to say, as Kantians say, that the agent does not set and pursue her own ends. Sarah Buss summarizes the challenge quite aptly:

> whether an instance of practical reasoning is self-determined is a matter of whether it is really the agent herself who is doing the reasoning.[15] And this would seem to depend on whether *she* determines her response to the considerations that figure in her reasoning […].

But when an agent acts under volitional coercion (or deception, which is what Buss focuses on), she is still the one doing her own practical reasoning.

As an illustration, consider the following discussion by Epictetus, who was himself born into slavery:

> It is reasonable to one man to hold another's chamber-pot for him, since he considers only that if he does not submit to this, he will be beaten and lose his dinner, whereas if he does hold it, he will have nothing harsh or distressing to suffer; whilst to some other man it appears insupportable, not merely to hold the pot himself, but to allow anyone else to do so. If you ask me, then, "Shall I hold the pot or not?", I will tell you, it is a more valuable thing to get a dinner than not; and a greater disgrace to be given a thrashing than not to be: so that, if you measure yourself by these things, go off and hold the pot.
>
> "Yes, but that would be beneath me."
>
> It is you who are to consider that, not I: for it is you who know yourself, and what value you set upon yourself, and at what rate you sell yourself: for different people sell themselves at different prices.[16]

What this passage brings out nicely is just how much agency is exercised even by someone acting under coercion. The agent still makes a choice; she still thinks about what to do, and does what she thinks will best advance her own ends in the

---

[15] Sarah Buss, "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints," *Ethics* 115 (2005): 214.
[16] Epictetus, *The Discourses of Epictetus*, trans. Elizabeth Carter and Robin Hard (London: Everyman, 1995), Book I, Chapter 2.

circumstances. Indeed, she could not abdicate the power and responsibility of choice even if she wanted to. In this sense, agency is fundamentally inalienable.

Kantians have tended to avoid recognizing this truth about agency. Sometimes, they suggest that the coerced agent pursues no ends in acting as she does. O'Neill, for instance, concedes that "[v]ictims [of coercion] may *want* the same ends as their coercers," but insists that "that is not the same thing as sharing those ends, for one who is coerced, even if pointlessly, is not pursuing, nor therefore sharing, ends at all."[17] But while this claim is more or less true of cases of mere coerced movement, it is obviously false of cases of coerced action proper. If a parent pays a ransom to free his kidnapped children, he is doing what he can to secure the safety of his children—an end which he pursued before the kidnapping, and which he will continue to pursue after it.

At other times, Kantians concede that the coerced agent exercises her agency or practical reason in some sense, but insist that what she is deprived of is the ability to determine her own ends, or alternatively that she is restricted to the pursuit of only a narrow, short-term range of goods such as immediate survival or the avoidance of pain.[18] This is false, too. When a prisoner signs a confession at gunpoint, her end of survival is by no means forced upon her by her captor. Even at gunpoint, she can deliberate whether it really is best to survive so that she can continue pursuing her various ends or whether the dishonor of yielding to the threat would render her survival worthless. The further ends in the pursuit of which she does this deliberation,

---

[17] O'Neill, "Between Consenting Adults," 262.
[18] David Zimmerman, who is sympathetic to the Kantian approach, makes both points in his "Coercive Wage Offers," ibid.10, no. 2 (1981): 130. See also Ripstein in the passages quoted above.

moreover, may be as grand and long-term as any: the quest to realize a communist utopia, say, or God's Kingdom on Earth.

2.2

The difficulty of explaining the unfreedom of coerced action also makes it difficult for Kantians to explain what is morally wrong with coercion.

Coercion, on the Kantian view, is supposed to involve some way of treating the other *as if* she were a mere thing rather than as a person. But the victim of this treatment is not actually a mere thing, and treating her *as* a thing does not turn her into one. We have seen how this makes it difficult to understand the unfreedom of coerced action. If freedom is a matter of being a person rather than a thing, then even the slave is free. But in that case, it equally becomes difficult to see what could be so very wrong with coercing another. For if the other's status as a practical reasoner is not actually threatened by volitional coercion, in what respect is she injured and wronged?

One might try suggesting that the wrong of coercion consists in something on the order of an insult. After all, the coercee's agency does seem to figure as a mere instrument or "machine" in the *coercer's* deliberation: rather than presenting the other with what he takes to be good reasons, he simply "tries to determine what levers to pull to get the desired results from [her]."[19] This way of viewing another person seems to amount to an insult to her dignity, a failure to acknowledge her standing as a practical reasoner.

---

[19] Korsgaard, "The Right to Lie," 141. She is talking about deception, but the point might equally apply to volitional coercion.

Now, first of all, the *main* problem with coercion is surely not that it has a certain symbolic or expressive significance. Indeed, Kant himself plausibly distinguishes "duties of respect" against arrogance, defamation, and ridicule—which he does seem to think are ways of merely *expressing* the denial of another's agency—from the duty against coercion.[20] But in any case, it is unclear what it is that the coercer fails—even symbolically—to recognize or respect about his victim's status as a practical reasoner. After all, the coercer knows, and even counts on the fact, that the coercee will respond to his threat by doing her own practical reasoning in the pursuit of her own ends. He expects his victim, in other words, to exercise her agency. So where is the insult or failure of acknowledgment?

Kantians speak in very evocative and compelling terms of coercing another as using her agency in a merely "causal" way, in the manner of a mere "machine" or "instrument." But it is not enough just to say such things; one has to make them intelligible. In particular, the key words here need to be given a determinate content which is compatible with the fundamental inalienability of agency. In the meantime, the Kantian picture seems to raise more questions than it answers. What it lacks, specifically, is an explanation of the sense in which an agent's practical reason can be impeded or subverted without being suspended or destroyed altogether.

3. MORALISM

One might respond to these challenges by seeking an explanation of what is wrong with

---

[20] See Japa Pallikkathayil, "Deriving Morality from Politics: Rethinking the Formula of Humanity," *Ethics* 121, no. 1 (2010): 141–142.

coercion which is not based on a prior account of autonomy. The most important version of this approach is what I call moralism. Though it is favored by many Kantians, it amounts to a considerable departure from Kantianism properly speaking, since it displaces the concept of autonomy from its traditional place at the foundation of the ethics of interpersonal relations.

Moralism of one or another sort is defended by many philosophers including Robert Nozick (sometimes), Alan Wertheimer, T.M. Scanlon, Arthur Ripstein, Japa Pallikkathayil, and others.[21] While there are many differences between different versions of the approach, it exhibits a certain unity of both method and content. Methodologically, the moralists propose to begin not with the question why coerced action is unfree but with the explicitly moral question why to coerce another is to wrong her. Substantively, their proposal is that coercing another wrongs her because it involves violating some independent right of hers.

If our first interest is in the moral question, it is natural to begin with an observation along the following lines. Coercion constrains its victim's options, attaching consequences to the non-performance of an action which make it undesirable or unreasonable for the victim not to comply—but not all transactions premised on such

---

[21] Nozick, *Anarchy, State, and Utopia*, 262–264; Alan Wertheimer, *Coercion* (Princeton: Princeton University Press, 1987); Scanlon, *Moral Dimensions*, Chapter 2; Ripstein, *Force and Freedom*; Japa Pallikkathayil, "The Possibility of Choice: Three Accounts of the Problem with Coercion," *Philosophers' Imprint* 11, no. 16 (2011). See also, e.g., Frank Knight, *Freedom and Reform* (New York: Harper & Brothers, 1947); Vinit Haksar, "Coercive Proposals [Rawls and Gandhi]," *Political Theory* 4, no. 1 (1976); Benjamin Sachs, "Why Coercion is Wrong When It's Wrong," *Australasian Journal of Philosophy* 91, no. 1 (2013); Saba Bazargan, "Moral Coercion," *Philosophers' Imprint* 14, no. 11 (2014); and Stephen J. White, "On the Moral Objection to Coercion," *Philosophy and Public Affairs* 45, no. 3 (2017).

constraints appear to be wrongful. For example, Pallikkathayil invites us to contrast the case of "the mugger who says, 'Your money or your life', meaning, 'I will kill you if and only if you refuse to hand over your money'"—a paradigm case of coercion—"with the following lunchtime interaction. I say to you, 'I will give you my pretzels if (and only if) you give me your chips'."[22]

According to the most common version of the moralistic approach, what sets the coercer apart is that he characteristically proposes to do something independently permissible to his victim.[23] The mugger, for instance, has no right to take his victim's life, and would violate his victim's rights if he did so; he therefore wrongs his victim merely by threatening to do so, and since his constraint of his victim's options is premised on this threat, it is impermissible and a way of wronging her. By contrast, when I propose to withhold my pretzels from you unless you give me your chips, I do not wrong you, since you have no entitlement to my pretzels and I would be acting within my rights in withholding them. As Ripstein puts it, the coercer, unlike the mere contractor, offers his victim "something that he has no right to offer."[24]

In this way, moralists seem to be able to account for the wrongfulness of coercion. On the most common version of the view, to coerce another is to wrong her, in the final analysis, because it involves threatening to wrong her: "the wrong of coercion [is] parasitic on the wrongfulness of acting on the intention being announced."[25] This explanation does not require us to deny the fundamental inalienability of agency. For

---

[22] Pallikkathayil, "The Possibility of Choice," 1, 2.
[23] Ibid., 10ff.
[24] Ripstein, *Force and Freedom*, 132.
[25] Pallikkathayil, "The Possibility of Choice," 18.

the problem with coercion is not that it deprives another of the power of choice or otherwise subverts or bypasses her agency, but rather that it involves an independent violation of her rights.

There are various modifications of this basic approach. Ben Sachs considers it necessary to explain more substantially why the speech act of threatening to do something impermissible is itself impermissible; he argues that such a speech act is wrong because it makes one more likely to carry out the threatened impermissible action.[26] Saba Bazargan argues that *in addition* to being wrong because it involves a threat to violate someone's rights, coercion is wrong because it involves "hacking" someone's aims: making it the case that she would be in a better position to achieve those ends if she did not have them, so that the coercer could not hold them hostage.[27]

Stephen White drops the requirement that the threatened action be impermissible altogether. He explains the wrongfulness of coercion in terms of the coercer's wrongfully making his victim responsible for ensuring that the coercer does not carry out his threat.[28] For instance, if Green tells Brown, "Stay out of Malibu, or you'll be sorry," Green thereby transfers to Brown the responsibility of ensuring that she does not beat him up—but this responsibility ought to belong to Green, not to Brown. Here, it is not necessarily the threatened act itself but rather the transfer of

---

[26] Sachs, "Why Coercion is Wrong When It's Wrong," I am reminded irresistibly of this passage from Anscombe: "One might not have a 'mind' to do something, distinguishable from uttering the words. And then, as Quine once put it (at a philosophical meeting), one might do the thing 'to make an honest proposition' of what one had said." G. E. M. Anscombe, *Intention* (Cambridge, MA: Harvard University Press, 1963), 4.
[27] Bazargan, "Moral Coercion."
[28] White, "On the Moral Objection to Coercion."

responsibility which is independently impermissible. What all of these accounts have in common, however, is that they explain the wrongfulness of coercion not on the basis of the unfreedom of coerced action but on the basis of the violation of some independent claim.

4. The priority of freedom to right

Moralism faces a number of serious difficulties, all of which boil down to its defining structural feature, which is the way in which it attempts to understand the wrongfulness of coercion as essentially independent of the coercee's unfreedom. In fact, we cannot properly bring into view the distinctive violation of right which coercion involves without first understanding the way in which it undermines the coercee's freedom.

4.1

Moralism does not promise us a satisfactory account of the unfreedom of coerced action. The problem is that it is hard to shake the naïve thought that the freedom or unfreedom of an action must ultimately be a matter of the way in which the action expresses or fails to express its agent's will. And from this point of view, the concept of rights simply seems to be neither here nor there. Why should I only count as acting involuntarily if a threat violates my rights? What we want is some understanding of how being coerced does not merely wrong one but how it undermines one's agency or practical reason. Otherwise, we still lack an explanation of how coercion makes its

victim, specifically, *unfree*.[29]

Now, some moralists take themselves to be able to give such an explanation. White proposes an explanation, in particular, of why the coercer's threat is not merely another feature of her victim's circumstances "with regard to which her victim is left free to respond as he sees fit."[30] For a coercer, White points out, cannot "legitimately" take her victim's situation to be one in which the victim is responsible for ensuring that the coercer does not carry out the wrongful threat; she cannot view her victim's situation in this way "in good faith."[31] He concludes:

> Looked at from Green's point of view, then, her threat can only be seen as interfering with Brown's ability to respond to his circumstances on his own terms. And because it is, in this sense, an interference with Brown's will aimed at ensuring he acts as Green intends for him to act, we can put our objection to the threat by saying that it amounts to an effort on Green's part to subject Brown to her will.[32]

White's explanation is unconvincing in two ways. First, from the fact that Green cannot "legitimately" view Brown's situation in a certain way it does not follow that she cannot view it in this way, period, or even that she cannot view it in this way "in good faith." If Green is the mafia boss she sounds like, she might mistakenly but sincerely take herself to have the prerogative to effect the relevant transfer of responsibility. The substance of White's claim, then, is really that Green would be making a *moral* mistake in taking herself to be offering Brown a reason for action on a par with any other. But it is at least not obvious what this tells us about Brown's unfreedom.

---

[29] For a version of this complaint, see Zimmerman, "Coercive Wage Offers."
[30] White, "On the Moral Objection to Coercion," 230.
[31] Ibid.
[32] Ibid.

Second, more importantly, the proposed explanation is limited essentially to the coercer's point of view. Indeed, White explicitly affirms that "from [the coercee's] point of view, the threat looks like a change in the coercee's situation to which she must now (freely) decide how to respond."[33] But this is a fatal concession. It leaves us in the same position as the original Kantian approach. How can we claim to understand the unfreedom of coerced action if we can discern no way in which such action is actually a privation of agency? How can we characterize Green's coercive act as "an effort […] to subject Brown to her will" if this effort is bound to fail because Brown will choose freely regardless? Admittedly, White says only that it "looks" like Brown chooses freely, and only from her own point of view. But why should the coercer's point of view be privileged in this situation? Why not the agent's own point of view? The problem is not epistemic, after all. In any case, the asymmetry of the account is confusing *per se*. An agent's freedom or unfreedom is not relative to a point of view. Either she chooses freely, or she does not.

Other moralists concede that they are not in the best position to solve the problem of unfreedom. Wertheimer, for instance, does make a proposal for how to explain the unfreedom of coerced action (which I will consider in Section 6), but he has evident reservations about it, conceding that moralism "preserves the moral force of voluntariness claims" but "threatens to preserve that moral force by detaching voluntariness from the will, from its psychological referents."[34] The reason he is not unduly troubled, though, is that it is the moral significance of coercion which primarily

---

[33] Ibid., 231.
[34] Wertheimer, *Coercion*, 305.

gives the concept its philosophical interest. The point, then, is that once we have explained the moral issue, we have explained what really matters.[35]

Now, it does seem to me that the inability to account for the unfreedom of coerced action is a decisive defect in a theory of coercion. The unfreedom of coerced action is too obvious to be ignored. Indeed, as we saw in Chapter I, there is a broad tradition of thought according to which coercion is *the* source of unfreedom *par excellence*.[36] That said, given the difficulties we encountered in trying to understand the issue of unfreedom, we should at least take seriously the proposal that there is philosophical progress to be made by focusing our efforts elsewhere. Let us see, then, how the moralistic account fares on its own terms, that is, purely as an account of the wrongfulness of coercion.

4.2

In fact, the failure to explain the unfreedom of coerced action ends up infecting the proposed explanation of the wrongfulness of coercion. Let me put the point in general terms first. Coercion is not just any old way of wronging someone, but specifically a way of wronging someone by making them unfree. According to moralism, coercion is wrongful not because it makes someone unfree but because it involves threatening to do something which is wrong on independent grounds. The account therefore explains the wrongfulness of coercion in the wrong way. While this is a completely general problem, we can isolate it and see it most vividly in those cases of coercion where the coercer's

---

[35] Compare Buss, "Valuing Autonomy and Respecting Persons," 226ff.
[36] Naturally, moralism also faces difficulties accounting for the non-interpersonal dimension of freedom and unfreedom—but I will leave this issue aside here.

threat is not independently wrong at all.[37]

Perhaps the clearest example is blackmail. Suppose a journalist discovers that a candidate for public office published some embarrassing or otherwise damaging material on the internet as a teenager. Presumably, the journalist has the right to make this information public, just as he has the right to keep it to himself. Presumably he likewise has the right to tell the candidate of his plan to publish the information. If, however, he approaches the candidate and threatens to publish the material *unless* she pays him or performs some other action, his proposal constitutes blackmail and wrongs her—or so it would seem. But on the standard moralistic understanding of coercion, this is difficult to explain, for the obvious reason that what the blackmailer threatens to do is not, as in the case of the master and slave, independently wrong. If anything, the blackmailer does his victim a favor by giving her the option of buying his silence, something to which she has no independent entitlement.

For a second example, consider a case of sexual harassment. Suppose that an employer tells his employee that he will fire her unless she has sex with him. Certainly as a matter of law in an employment-at-will context (as in the United States), the employee has no generic and independent right that she not be fired; we may even suppose, for the sake of argument, that the employee has previously exhibited incompetence or engaged in gross misconduct of a sort which would normally be reasonable grounds for firing her. The employer, then, may well have the right to fire

---

[37] While my criticism in its general form applies to moralism in general, what I say about these examples in particular applies only to the standard version of the approach and not, for instance, to White's version.

her, as well as the right to tell her that he plans to do so. And yet when he threatens to fire her *unless* she has sex with him, his proposal clearly becomes coercive and wrongs her. Again, on a moralistic understanding of coercion, this is difficult to explain, since what the employer threatens to do—namely to fire her—does not appear independently wrong. Indeed, theoretically, as in the case of blackmail, it could even be the case that a sufficiently desperate employee would rather receive her employer's "offer" than be fired outright. (Of course, she would prefer simply to keep her job without strings attached, but this option is not on the table, and we are supposing that she has no independent entitlement to it.)

Though moralists have difficulty dealing with these cases, there is of course a very straightforward explanation of why blackmail and *quid pro quo* sexual harassment are wrong: they are wrong because they are coercive, because they involve the subjection of one person's will to that of another. What such cases illustrate, in other words, is the general truth that coercion is not a parasitic wrong. It is wrong not merely on account of the independent wrongfulness of some other action which it involves but because of its coerciveness. Moralism explains the wrongfulness of coercion in the wrong way.

Now, it should be said that moralists have made some attempts to accommodate the difficult cases. In general, they have done so by redescribing or contextualizing the coercer's threatened action, or the act of the threat itself, in a way which reveals its "independent" wrongfulness. Wertheimer suggests that what is wrong with the threat involved in blackmail, for instance, might be "that it is wrong to assert [one's] rights to gain advantages with which those rights have no intrinsic connection and which they

are not designed to serve."[38] Regarding the sexual harassment case, Pallikkathayil suggests that the problem may boil down to the general objectionableness of employment-at-will, which effectively allows for "an abrupt and radical change in job description."[39] As with Wertheimer's suggestion, what is at issue appears to be the fact that the coercer puts his prerogatives to use for a purpose which is not intrinsically related to that prerogative, in this case because it is unrelated to the job description.

But why would some such constraint on the relation between the threat and the action which it is meant to induce be a moral requirement? Why would violating this constraint be a way of wronging someone? Scanlon gives us more to work with. Speaking about the case of sexual harassment, he likewise suggests that the action which the employer threatens and which would be impermissible is not simply *firing her* but *firing her because she did not comply with his threat*.[40] But he explains, further, that this action would be impermissible because if the employer were to have unrestricted discretion to fire his employees "for any reasons whatever" he would have "an unacceptable form of control over others."[41]

This is a telling moment. It seems to me that what we are finding here is that the most convincing moralistic proposals for dealing with these cases are convincing

---

[38] Wertheimer, *Coercion*, 220. He is really speaking about the coerciveness of threatening criminal prosecution in order to secure private agreements—but he adds that "[s]imilar considerations may motivate the criminalization of blackmail." Ibid.
[39] Pallikkathayil, "The Possibility of Choice," 18–19. She also points out that there may in fact be institutional or contractual constraints on the reasons for which the employer may permissibly fire his employee (18), and that "acting on the employer's conditional intention involves paying for sex," which "might be impermissible" (19).
[40] Scanlon, *Moral Dimensions*, 84.
[41] Ibid. Scanlon takes this account to apply to "abuses of privilege" in general (86).

precisely because they approximate the more straightforward, non-moralistic type of explanation. Scanlon's objection is that the employer's action is the exercise of an unacceptable form of control over others—but that is exactly what coercion is. So his explanation looks very much like this: the employer wrongs his employee because he coerces her. But then coercion is not a parasitic wrong, after all.

In any case, we will now want to know what "kind" of control is at issue, why it constitutes a form of "control," and why exactly it is "unacceptable." In fact, these are just the sorts of things which we would expect a theory of coercion to tell us. But the moralists are simply not well placed to give us these explanations. They give us proposals guided by more or less *ad hoc* moral reasoning about these cases, *instead* of by a general account of some form of relation to another's will which is wrong in itself, wrong because it is incompatible with the other's freedom. This shows, I think, how far behind we have left some of the core principles of Kant's moral philosophy.

4.3

This brings me to my final, related complaint against moralism. What happens when we chase the moralistic chain of moral explanation back to its foundation? The wrongfulness of coercion is supposed to be explained in terms of the independent wrongfulness of some act which it involves. What explains the wrongfulness of this independently wrongful act? In the case of the sexual harasser, we had reason to think that what explains the wrongfulness of the "independently" wrongful act was in fact its coerciveness—and this presented some problems for the moralistic approach. But I think this difficulty is likely to be quite general, on account of the fundamental place which the concepts of freedom and coercion tend to occupy in our moral thought and

especially in our thought about rights.

Even if we confine our attention to more central cases of coercion which do not involve blackmail or the like, we are entitled to ask what ultimately underwrites the rights whose violation is supposed to explain the wrongfulness of coercive acts. Moralists cannot at this point appeal to a prior concept of freedom or coercion, since they take the wrong of coercion to be parasitic on some *other* wrong. But this constraint threatens to place an unsustainable theoretical burden on the concept of right.

Consider the mugging. Why would the mugger be wronging his victim if he shot her? One could give two very different kinds of answers. On the one hand, one might appeal to something on the order of an especially pressing *interest* or *end* of the victim, by virtue of which shooting her constitutes an especially egregious *infliction of harm*. On the other hand, one might appeal instead to the special status of the body as a condition of the possibility of freedom—and, more fundamentally, to a right to freedom and against coercion.

It is arguably the latter, more formal sort of thought which animates the tradition of thought about rights. In Kant's philosophy of right, for instance, the right to freedom has a foundational status:

> *Freedom* (independence from being constrained by another's choice), insofar as it can coexist with the freedom of every other in accordance with a universal law, is the only original right belonging to every man by virtue of his humanity.[42]

Indeed, Kant defines the very concept of right in terms of freedom:

---

[42] Immanuel Kant, *The Metaphysics of Morals*, in *Practical Philosophy*, ed. and trans. Mary Gregor (Cambridge: Cambridge University Press, 1996), 6:237, emphasis his.

> Any action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law.[43]

For what it is worth, this idea is also in keeping with the spirit of Kant's moral (as opposed to political) philosophy. To quote Korsgaard: "According to the Formula of Humanity, coercion and deception are the most fundamental forms of wrongdoing to others – the roots of all evil."[44] There is a powerful philosophical rationale for this formal kind of approach to the philosophy of right. The point is that the mugging victim's right against assault does not depend fundamentally on her contingent preference not to be assaulted. It is an entitlement which she has just by virtue of her status as an agent, and because of the way security against assault is a necessary condition of free agency.

One finds these sorts of ideas, in fact, in the writings of some moralists, such as Ripstein.[45] But it is quite unclear how moralists can be entitled to them. Assuming, as Ripstein does, that freedom is what one has when one's will is not subject to another's, and assuming, as he also does, that being subject to another's will is a matter of having some independent right violated, then Kant's idea of a *fundamental* right to freedom is literally nonsense. Such a right could not be fundamental because its violation is essentially parasitic on the violation of some other, independent rights. It seems to me, then, that moralists are at risk of retaining a coherent concept of rights only at the cost of undermining at least one key motivation for employing such a concept in the first place.

---

[43] Ibid., 6:230, emphasis his.
[44] Korsgaard, "The Right to Lie," 140.
[45] Ripstein, *Force and Freedom*.

Taking a step back, reflection both on central cases such as the mugging and on more complicated cases such as blackmail and *quid pro quo* sexual harassment strongly suggests that the concept of the relevant kind of right is posterior to the concept of coercion, and that the distinctive wrongfulness of coercive acts is posterior to their coerciveness. But if this is correct, then the moralistic attempt to explain the wrongfulness of coercion without relying on an account of autonomy cannot succeed, for it reverses the proper order of explanation. We cannot evade the challenge of making sense of autonomy and the way coercion undermines it; we must address that challenge head-on.

5. PSYCHOLOGISM

The final major approach to thinking about freedom and unfreedom is what I call psychologism. Like Kantianism, it begins with a conception of autonomy, but its foundational concept is not practical reason but desire. And it explains the unfreedom of coerced action in particular by appeal to a distinction between first-order and higher-order desires.

5.1

The starting point of psychologism, as of Kantianism, is the idea that free activity is determined by and an expression of its agent's will. But whereas Kant understands the will in the first instance as a power of practical reason, psychologists think of it in the first instance as a power of desire. As a first approximation, then, acting freely is a

matter of (non-accidentally) doing what one wants to do.[46] If I stay at home because I have decided to spend a quiet evening with my family, my action expresses my will; I do what I want to do, so I do it freely. On the other hand, if I would really like to go outside but I stay at home because someone has locked me in, my action reflects not my desire but merely an alien force to which I am subject, in this case another's will; I do not do what I want to do, and my "action" is thus involuntary, unfree.

The modern "home tradition" of psychologism is British empiricism. According to Thomas Hobbes, "he is free to do a thing, that may do it if he have the will to do it, and may forbear if he have the will to forbear."[47] John Locke expresses much the same view: "so far as a Man has a power to think, or not to think; to move, or not to move, according to the preference or direction of his own mind, so far is a Man *Free*."[48] And so does David Hume: "By liberty, then, we can only mean a power of acting or not acting, according to the determinations of the will; that is, if we choose to remain at rest, we may; if we choose to move, we also may."[49] Most famously, John Stuart Mill says that "liberty consists in doing what one desires" (though is considered view is more sophisticated).[50]

---

[46] I will not analyze the non-accidentality condition further, as it is not relevant to my argument here.
[47] Thomas Hobbes, "Of Liberty and Necessity," in *Hobbes and Bramhall on Liberty and Necessity*, ed. Vere Chappell (Cambridge: Cambridge University Press, 1999), 16.
[48] John Locke, *An Essay Concerning Human Understanding* (Oxford: Oxford University Press, 1975), Book II, Chapter XXI, Section 21. He makes similar statements throughout the same chapter. E.g.: "Liberty, 'tis plain, consists in a power to do or not to do; to do or forbear doing as we will. This cannot be denied." Ibid., Section 56.
[49] David Hume, *An Enquiry Concerning Human Understanding*, Second ed. (Indianapolis: Hackett, 2011), Section VIII, 95.
[50] John Stuart Mill, *On Liberty*, in *On Liberty and The Subjection of Women*, ed. Alan Ryan (London: Penguin, 2006), Chapter V, 109. For the nuances, see ibid., Chapter III, 70; and

This simple version of the view, though, is too simple. For it is possible to be alienated from one's own desires. It is possible to have, and act from, a desire which one experiences not as genuinely or fully one's own but as the manifestation of an alien compulsion. For instance, I might stay at home neither because I have decided to spend a quiet evening with my family nor because someone has locked me in, but because I have a severe case of agoraphobia and the prospect of going outside and encountering crowds of people seems intolerable to me. I may experience this phobia not as an expression of my true will but as a debilitating impediment, an obstacle to my freedom. In this case, my action expresses my desire—for my phobia does issue in a desire to stay at home—yet it does not express my will; it is not truly free. More precisely, then, doing what one wants to do is a necessary but not a sufficient condition of freedom.

In the twentieth century, philosophers such John Plamenatz, Gerald Dworkin, and Harry Frankfurt have articulated a further necessary condition of freedom in terms of the notion of higher-order desires: desires regarding our desires.[51] The idea is that we not only want to do and refrain from various actions; we also want to have and lack various desires, or at least to want various desires to issue in action or to remain inert. When I experience my agoraphobia as an oppressive compulsion, for instance, I have a desire to stay indoors but I wish I did not have this desire, or at least that it did not decisively move me to stay indoors. And this explains my unfreedom. For when I act

---

John Stuart Mill, *The Subjection of Women*, in *On Liberty and The Subjection of Women*, ed. Alan Ryan (London: Penguin, 2006), Chapter I, 146–149.

[51] J. P. Plamenatz, *Consent, Freedom and Political Obligation* (London: Oxford University Press, 1938), Chapter V; Gerald Dworkin, "Acting Freely," *Noûs* 4, no. 4 (1970); Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68, no. 1 (1971).

from a motive which I wish I did not have, there is a sense in which I act "against my own will," even though I do what I want to do.

There are two ways in which the second necessary condition of freedom is spelled out. Sometimes it is argued that freedom requires positive endorsement of the first-order desires from which one acts: it is a matter of (non-accidentally) wanting *what one wants to want*.[52] Sometimes it is merely argued, more weakly, that freedom requires the absence of disavowal or resentment of the first-order desires from which one acts: it is a matter of (non-accidentally) wanting *what one does not want not to want*.[53] Which version we pick makes no difference for my purposes, though I will mostly focus on the weaker version, since this is usually invoked in the context of discussion of coercion in particular.

Either way, we could call this higher-order condition the "internal" condition of freedom. It regulates the constitution of the agent's will, understood in terms of her desires. The sort of unfreedom which violates this condition will necessarily amount to some kind of problem internal to the agent's desires (even if it is *caused* by something external to the will). The first-order condition, by the same token, could be called the "external" condition, since it regulates the outward expression of the agent's will in action. The sort of unfreedom which violates this condition will not consist in a problem internal to the agent's desires, which may be perfectly well formed as far as this particular form of unfreedom is concerned. For the same reasons, Frankfurt suggests that the external condition corresponds to "freedom of action," while the internal one

---

[52] E.g., "Freedom of the Will and the Concept of a Person."
[53] E.g., Plamenatz, *Consent, Freedom and Political Obligation*; Dworkin, "Acting Freely."

corresponds to "freedom of the will."[54]

This basic framework can be filled out and modified in various ways. Dworkin himself, for example, amends his view in a later book, where he defines autonomy as a more dynamic capacity to reflect on and change one's first-order preferences in light of one's higher-order preferences.[55] For the most part, such differences do not concern me here. What I take to be definitive of psychologism is the idea that freedom consists in (non-accidental) coherence among higher and lower orders of desire as well as the (non-accidental) expression of desire in action, where desire is understood as a subjective psychological state. Critiques and amendments which reinterpret the internal condition of freedom in an objectivist way (such as those associated with Gary Watson, Susan Wolf, and Charles Taylor) therefore represent significant departures from psychologism, and I will deal with them separately at the end of the next section.

5.2

How can a psychologistic theory of autonomy help us make sense of coercion?

The account of direct physical coercion would be fairly straightforward. If someone grabs my body to use as a human shield against deadly projectiles, it is not the case that I move by body into harm's way because I want to. Or again, if I stay indoors

---

[54] Frankfurt, "Freedom of the Will and the Concept of a Person," 136.
[55] Gerald Dworkin, *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press, 1988). For more recent work in this tradition, see, e.g., Michael E. Bratman, *Faces of Intention: Selected Essays on Intention and Agency* (Cambridge: Cambridge University Press, 1999), Chapter 10; J. David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000); Suzy Killmister, "The Woody Allen Puzzle: How 'Authentic Alienation' Complicates Autonomy," *Noûs* 49, no. 4 (2015); and *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance* (Routledge, 2018).

because someone has locked me in, I do not stay indoors because I want to. So the external, first-order condition alone is sufficient for explaining the unfreedom of these "actions."

What about cases of coercion mediated in some way by the coerced agent's own desire? The internal, higher-order condition allows us to deal in an obvious way with some of these cases, namely those which involve what the law calls an "overborne will." An agent may divulge a secret under torture in order to avoid further suffering, while wishing that she had the inner strength to take the secret to her grave. This agent suffers what Frankfurt calls an "inner defeat."[56] She acts "against her own will" because she acts from a motive she wishes she did not have, or at any rate a motive which she wishes were inefficacious on this occasion. She therefore acts unfreely.

Not all coercion, however, involves an overborne will. The victim of a mugging, for instance, presumably does not normally suffer an "inner defeat" like the victim of torture imagined above. She does not wish that she had the strength of will to take the money to her grave.[57] Given the circumstances, she not only wants to hand over the money; she wants to have this desire and to act from it. Her first-order desire is not an object of her resentment. What, then, is the "motive" which she resents? How can we make sense of the unfreedom of coerced action without an overborne will?

Typically, psychologists propose to get around this difficulty by emphasizing not the coerced agent's immediate first-order desire to perform the coerced action but the

---

[56] Harry Frankfurt, "Coercion and Moral Responsibility," in *Essays on Freedom of Action*, ed. Ted Honderich (London: Routledge & Kegan Paul, 1973), 81.
[57] Frankfurt acknowledges this difference. Ibid., 81–82.

further end for the sake of which she performs this action. Frankfurt, taking a representative line, suggests that

> a person who submits to (and who does not merely [accidentally] comply with) a threat necessarily does so in order to avoid a penalty. That is, his motive is not to improve his condition but to keep it from becoming worse. This seems sufficient to account for the fact that he would prefer to have a different motive for acting.[58]

Dworkin says something very similar:

> Men resent acting for certain reasons; they would not choose to be motivated in certain ways. They mind acting simply in order to preserve a present level of welfare against diminution by another. They resent acting simply in order to avoid unpleasant consequences with no attendant promotion of their own interests and welfare.[59]

But this is still incredibly quick. The proposal cannot simply be to identify the further end with which the coerced agent acts as the object of resentment. For in the same sense in which the mugger's victim does not resent her first-order desire to hand over her money (in the circumstances), she presumably does not resent her further desire to preserve her life. In this respect, her case differs once again from the case of the victim of torture, who may well resent her further desire to avoid further pain in this context.

Instead, what the psychologists have in mind must be something like the following. Assuming, say, that the coerced agent hands over the money in order to preserve her life, she resents neither her desire to hand over the money nor her desire to

---

[58] Ibid., 82–83. He adds a further condition to the effect that the coerced person must be moved by a first-order desire which is literally "irresistible," but this assimilates coercion in general implausibly to the particular case of coercion involving an overborne will. Other psychologistic philosophers do not insist on the irresistibility condition.

[59] Dworkin, "Acting Freely," 377. See also Plamenatz, *Consent, Freedom and Political Obligation*, 125; and Joan McGregor, "Bargaining Advantages and Coercion in the Market," *Philosophy Research Archives* 14 (1988–1989): 49n21.

preserve her life, but rather *having to hand over the money in order to preserve her life*. In other words, whereas the "motive" which the victim of an "inner defeat" resents is simply a first-order desire, the "motive" which the victim of standard coercion (without an overborne will) resents is not a first-order desire but a complex connection between two first-order desires (the desire to hand over the money and the desire to preserve her life).

This is a vexing complication, for it destroys the attractive simplicity of the theory. The word 'motive' has quietly come to be used in two different senses: to refer to a first-order desire on the one hand and a nexus between two first-order desires on the other hand. By the same token, 'resentment' now seems to have two senses: for it is hard to accept that resentment of a nexus between two desires is fundamentally the same kind of attitude as resentment of a desire *simpliciter*. All of this looks suspiciously *ad hoc*. I will revisit the issue later. For now, though, let us take it in our stride and move on.

6. THE PRIORITY OF FREEDOM TO RESENTMENT

The most obvious problem with psychologism is its subjectivism. In addition to obscuring some forms of internal unfreedom, this subjectivism undermines the explanation of the unfreedom of coerced action and makes it hard to explain the wrongfulness of coercion or to identify what is special about coercion vis-à-vis other forms of unfreedom. On the other hand, if we jettison the subjectivism while keeping the rest of the psychologistic approach, we simply lapse back into some form of Kantianism or moralism, without having made real progress toward solving the

difficulties of those approaches.

6.1

Psychologism has been criticized on account of the subjectivism of the internal, higher-order condition—that is, its ultimate dependence on whatever the agent's (higher-order) attitudes happen to be, rather than on some standard of rationality or goodness which is not derivative of those attitudes.[60] People's higher-order attitudes can be idiosyncratic and arbitrary, and can in general fail to track unfreedom in two ways.

On the one hand, an agent can act from a desire which she endorses and yet be acting from a compulsion that makes her unfree. Consider, for example, the phenomenon of maladaptive preferences, where an agent forms irrational or otherwise suboptimal desires in response to entrenched, seemingly inescapable deprivation or oppression.[61] For instance, someone might form a genuine desire to stay in an abusive relationship, but only because she has come to believe that there is no escape from her situation. (Of course, that is unlikely to be the whole story—but it could be part of the

---

[60] See, e.g., Susan Wolf, "Asymmetrical Freedom," *The Journal of Philosophy* 77, no. 3 (1980); Charles Taylor, "What's Wrong with Negative Liberty," in *Philosophical Papers* (Cambridge: Cambridge University Press, 1985). See also Gary Watson, "Free Agency," *The Journal of Philosophy* 72, no. 8 (1975) for a related criticism.

[61] See Amartya Sen, "Justice: Means vs. Freedoms," *Philosophy and Public Affairs* 19, no. 2 (1990); *Commodities and Capabilities* (New Delhi: Oxford University Press, 1999); Martha Nussbaum, *Women and Human Development: The Capabilities Approach* (Cambridge: Cambridge University Press, 2000); and "Symposium on Amartya Sen's Philosophy: 5. Adaptive Preferences and Women's Options," *Economics and Philosophy* 17, no. 1 (2001). Sen's and Nussbaum's accounts of maladaptive preferences are especially pertinent because of their appeal to a substantive normative standard for the appraisal of desires. Contrast, e.g., the more formal accounts of Jon Elster, *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge: Cambridge University Press, 1983) and Luc Bovens, "Sour Grapes and Character Planning," *The Journal of Philosophy* 89 (1992).

story.) This person might well come to "identify" with her revised desire. Yet she remains unfree. The desire which motivates her is more like an alien compulsion than like the expression of a free will. The problem is not simply that some part of her is "out of sync" with the rest; it is that her desire is distorted in light of a more objective ethical or epistemic standard.

On the other hand, it also seems possible for people to resent the way they are motivated when they are not in fact unfree. This might on occasion be true, for instance, of someone with a wildly exaggerated and unreasonable sense of entitlement. These two possibilities are two sides of the same coin. The general problem is that because of its subjectivism, psychologism is unable to make sense of the difference between *being unfree* and *merely feeling unfree*.

Now, this problem fatally undermines the psychologistic account of coercion. First, it directly undermines the explanation of the *unfreedom* of coerced action. The idea is supposed to be that we are unfree when we are coerced because we resent the sort of motive from which we have to act when we are coerced. But if people can feel such resentment when they are not unfree, and fail to feel it when they are unfree, then this resentment cannot possibly be the ground of their unfreedom.

Second, and relatedly, the subjectivism makes it hard to explain the *wrongfulness* of coercion. For if an agent's freedom or unfreedom is so dependent on their subjective attitudes, it becomes very unclear why coercion should have the sort of moral significance that we take it to have. Specifically, we take it that we generally have at least defeasible second-personal claims against other people that they not coerce us. But if the unfreedom of coerced action is *merely* a matter of some attitude of the agent

toward it, it is not clear that we could reasonably make such claims. To be sure, being made to do something against one's will in this sense might not be desired, and might even be *undesirable*, but how could it possibly amount to being *wronged*? It is hardly obvious that we have a general claim not to be subjected to any treatment which, as a matter of our subjective make-up, we happen to resent.[62] The upshot is that psychologism will fail to explain the distinctive moral significance of coercion.

A final and related defect of psychologism is its inability to account for the distinctiveness of coercion vis-a-vis other, non-interpersonal forms of necessitation or unfreedom. It is one thing to have to undergo kidney surgery to remove a diseased organ. It is quite another thing to undergo kidney surgery because one has been kidnapped and will be released only on the condition that one "donate" a kidney to the kidnapper, who is suffering from kidney disease. In both cases, one presumably counts as acting "not to improve one's condition but to keep it from becoming worse"—a motive which we are told is a normal object of resentment. Yet there is surely a world of difference between the two cases.[63]

Defenders of psychologism sometimes more or less shrug this problem off. Here is Frankfurt:

> We do tend, of course, to be more resentful when another person places obstacles in our way than when the environment does so. What accounts for this greater resentment is not, however, the love of liberty. It is pride;

---

[62] Moralists sometimes press this sort of objection. See, e.g., Wertheimer, *Coercion*, 305; and White, "On the Moral Objection to Coercion."
[63] The last two objections also apply to conceptions of freedom as the mere possession of a range of "options." See, e.g., Thomas Hurka, "Why Value Autonomy?," *Social Theory and Practice* 13, no. 3 (1987); Philippe Van Parijs, *Real Freedom for All: What (If Anything) Can Justify Capitalism?* (Oxford: Oxford University Press, 1995); and Ian Carter, *A Measure of Freedom* (Oxford: Oxford University Press, 1999).

> or, what is closely related to pride, a sense of injustice. Only another person can *coerce* us, or interfere with our *social* or *political* freedom, but this is no more than a matter of useful terminology.[64]

If there were no adequate alternative to psychologism, then we might well be tempted to accept this consequence. But it is certainly a major concession. Even if both cases of kidney surgery can be said to involve *some* kind of unfreedom—which seems plausible—there nonetheless appears to be a distinctive dimension in which we choose voluntarily and exercise our practical reason fully when we donate the kidney out of sheer medical necessity and altruism, even if we do not choose and exercise our practical reason completely without qualification. Psychologism thus fails to bring into view not only what is distinctively wrong (morally) with coercing someone, but also what is distinctively wrong (from the point of view of freedom) with being coerced.

6.2

Some philosophers have proposed to leave behind the subjectivism of psychologism by defending objectivist versions of the internal condition of freedom. Susan Wolf identifies freedom with "the freedom to be determined by the True and the Good," which she identifies in turn with the possession of a set of powers, "in particular, the abilities necessary to see and understand the reasons and interests [one] ought to see and understand and the abilities necessary to direct [one's] actions in accordance with these reasons and interests."[65] We could say, then, that acting freely requires not merely

---

[64] Frankfurt, "Coercion and Moral Responsibility," 83–84.
[65] Wolf, "Asymmetrical Freedom," 160, 166. See also *Freedom within Reason* (New York: Oxford University Press, 1990), e.g., 73; and "Sanity and the Metaphysics of Responsibility," in *Responsibility, Character and the Emotions*, ed. Ferdinand Schoeman (Cambridge: Cambridge University Press, 1988), e.g., 159.

(non-accidentally) doing what one wants to do, but also, roughly speaking, that one's desires be unimpeded exercises of a power of cognitive and ethical knowledge. Call this kind of view "rational psychologism."

This sort of view seems capable of dealing with the special problems of "inner" unfreedom that stumped psychologism proper—the case of maladaptive preferences, for instance. But how does it help us make sense of coercion? Presumably, in place of an actual attitude of resentment, we would point to an idealized version thereof. The rational or reasonable agent, we might say, resents acting from the sort of motives from which she has to act when she is coerced. Nozick defends a version of this account in his early article on coercion. Rather than seeing the unfreedom of coerced action as a matter of the actual coerced agent's higher-order desires, he speaks of the higher-order desires of "the Rational Man." Credible threats normally constrain our freedom, he suggests, because "the Rational Man will normally not welcome credible threats, will normally be unwilling to be threatened, even if he is able to resist going along with them."[66]

One difficulty with this proposal is familiar from debates about hypothetical consent: it is one thing to show that one would consent to some condition in certain idealized circumstances, and another to show that one actually has consented.[67] On the

---

[66] Robert Nozick, "Coercion," in *Philosophy, Science, and Method*, ed. S. Morgenbesser, P. Suppes, and M. White (New York: 1969), 460.

[67] We might try saying that agents generally have a desire of an even higher order to be guided by "the Rational Man's" higher-order desires. This is roughly the sort of solution favored by Wertheimer, who defends a moralized version of Nozick's account. Wertheimer, *Coercion*, 297ff. But the solution merely reintroduces the problem at a higher level, since an actual agent might equally turn out to lack the relevant highest-order desire.

original psychologistic approach, the agent's *actual* (higher-order) attitudes were meant to be the basis on which to explain the sense in which the coerced agent necessarily acts "against her will." It is not as obvious how the satisfaction or frustration of merely *hypothetical* desire might explain an agent's actual freedom or unfreedom. In other words, the fatal subjectivism of psychologism is not a straightforwardly dispensable feature of that theory; it seems essential to the very form of the psychologistic explanation of the unfreedom of coerced action.

Putting this difficulty aside, though, there is a more fundamental problem. For now that we are no longer dealing with actual agents and their idiosyncratic passions, we are entitled to ask *why* "the Rational Man" should distinctively resent being credibly threatened. That is, what makes coercion a reasonable object of resentment? To be sure, as Nozick and many others point out, coercion usually makes one *worse off* in some way (at least by worsening one's "choice situation" relative to some baseline). This might give us reason not to "welcome" it. However, it is too thin a basis for the distinctive kind of resentment which belongs to coerced action. There are many things which make us worse off relative to some baseline and which we might reasonably "not welcome" which nonetheless do not make us unfree, let alone make us unfree in a way which gives rise to moral claims against others to act or refrain from acting in certain ways. What is needed for unfreedom with the relevant moral significance is a more specific form of resentment. It is the kind of resentment which we have reason to feel when we are wronged.

To say that "the Rational Man" resents being coerced is really to say no more than that coercion wrongs the coercee, for it is this fact which makes the coercee's

resentment reasonable. The explanation of the unfreedom of coercion, then, is ultimately based on the independent wrongfulness of coercion—independent, that is, of its effect on the coercee's unfreedom, since it is supposed to explain this unfreedom. But now one gets a distinct sense of *déjà vu*. Rational psychologism turns out to be a version of moralism. As such, it suffers from the same defect, which is that it cannot explain what is wrong with coercion in the right way. What is wrong with coercion is that it is a way of subjecting someone's will—of making her unfree. But for the rational psychologist, as for the moralist, the coercee's unfreedom is not prior to but posterior to the wrongfulness of coercion. This gets things the wrong way around.

Psychologism proper, in its subjectivist guise, actually has the very same structural problem. Dworkin is quite explicit about the order of explanation which he proposes:

> Aristotle observes that "those who act under compulsion and unwillingly act with pain." I am arguing that this is a necessary fact. We only consider ourselves as being interfered with, as no longer acting on our own free will, when we find acting for certain reasons painful. To put the thesis epigrammatically; we do not find it painful to act because we are compelled; we consider ourselves compelled because we find it painful to act for these reasons.[68]

That is, for the (subjectivist) psychologist, the unfreedom of coerced action is parasitic on the psychological response to being coerced. But once we understand that this psychological response is a rational response to being wronged, we can see that the proposed explanation is no different, fundamentally, from the moralist explanation.

Moralism and psychologism give objective and subjective versions of one and

---

[68] Dworkin, "Acting Freely," 378–379.

the same account, and they make two versions of one and the same mistake. Both reverse the proper order of priority between the coerciveness of coercion and its wrongfulness. They confuse a consequence of unfreedom for its ground.

7. WHAT NOW?

Where do we go from here? I want to start by returning to the bump in the road which we hit while initially coming to grips with the psychologistic account of coercion.

In applying their account of autonomy to interpersonal coercion in particular, I pointed out that psychologists quietly introduce a new kind of "motive" to serve as the relevant object of higher-order resentment, a motive which is not a first-order desire but a nexus between first-order desires. For example, the victim of a mugging resents neither the desire *to preserve her life* nor, in the circumstances, the desire *to hand over the money*. If she could purge herself of either of these desires in the circumstances, she would not choose to do so—unlike, for instance, the victim of torture who suffers an "inner defeat" and wishes she had the strength to take her secret to the grave. What the victim of the mugging resents, it was suggested, is rather *having to hand over the money in order to preserve her life*, or perhaps more generically *having to do something merely in order to preserve her life*.[69]

Why does the account suddenly become so cumbersome, having to wheel in a whole new universe of "motives"—motives, moreover, whose conditions of identity

---

[69] For the sake of argument, I will just accept the claim that coercion involves these forms of motivation and that we do resent being motivated in these ways. But see my criticisms of the Kantians at the end of Section 2.1.

and individuation are suspiciously ill defined? I suggest that it is for the following reason. In the two-part psychologistic account of autonomy, the higher-order condition is designed to deal specifically with *internal* forms of unfreedom, which consist in some problem with the agent's own desires. But while *some* forms of coercion—namely those involving an overborne will, an "inner defeat"—fit this pattern, the paradigm cases do not. It is simply not accurate to describe a person who is coerced as resenting her own desires or motives. Her relevant first-order desires, and her ability to regulate these in the light of her higher-order desires, are perfectly in order. In this sense, internally, her *will* is not impaired. Her problem is not what Frankfurt calls "freedom of the will," but "freedom of action."

Interestingly, the traditional Kantian approach, in its own way, was scuppered by exactly the same difficulty. The Kantians, as we saw, are unable to account for the unfreedom of coerced action because in paradigm cases of volitional coercion, the coerced agent's power of practical reasoning seems to be working just fine. Internally, her *will* is not impaired—so her autonomy seems unharmed, shining like a jewel by its own light. It seems that as soon as we start talking about autonomy, we are irresistibly drawn toward accounts which can bring into view only *internal* failures of it. And then it seems that external unfreedom in relation to other people has nothing to do with autonomy after all, since "the slave in chains may well be 'autonomous'."[70] This is how

---

[70] Raymond Geuss, "Freedom as an Ideal," *Proceedings of the Aristotelian Society* Supp. Vol. 69 (1995): 92. See also Wertheimer, *Coercion*, 9–10; Ripstein, *Force and Freedom*, 15; Pallikkathayil, "The Possibility of Choice," 6; and more generally Berlin, "Two Concepts of Liberty."

we lose our grip on the unity of freedom.

But we have seen that the attempt to make sense of coercion without a prior account of autonomy is bound to fail. We must therefore resist the idea of autonomy as a purely internal self-relation. And I propose to do so by returning to the other part of psychologism—the external, first-order condition that acting freely requires doing what one wants to do. This condition, after all, governs not the formation of desire but its proper expression in action, and so seems to be just the sort of thing we are looking for. And is it not the most natural thing in the world to think that the victim of a mugging who hands over her money in order to save her life does not do what she wants to do? We should try to make sense of this thought.

CHAPTER III. FREEDOM AS ACTIVITY FOR ITS OWN SAKE

In the previous chapter, I argued that we need an account of autonomy in order to make sense of coercion, but I found extant accounts of autonomy wanting. While most of the philosophical debate around autonomy has concerned the nature of what I called the internal (higher-order) condition of freedom, I suggested that it is really the external (first-order) condition which should be expected to do the work of explaining external freedom and unfreedom, including coercion. In this chapter, I defend a partial account of autonomy based on a reinterpretation of this external condition of freedom, one which better captures what we care about insofar as we care about being free and living in a society of free people.

One does something freely, I argue, only if one does it for its own sake and not *merely* for the sake of further ends; I show that this account is compatible with the freedom of instrumental action, as long as such action is also done on account of some intrinsic value which is grounded in its instrumental value (Section 1). I sketch some of the implications of the account in order to give an initial impression of its explanatory power (Section 2). In the rest of the chapter, I refine the account in several ways. I argue that the unfreedom of merely instrumental action varies in degree and significance, and is especially egregious when the action is done for the sake of a large range of important ends (Section 3) and occupies an independently significant place in the agent's life (Section 4). I also argue that action done for the sake of pleasure, properly understood, is thereby done for its own sake and to that extent freely, but that action done from

moral duty is not thereby done for its own sake, and can be done unfreely (Section 5).

1. FREEDOM AND INTRINSIC VALUE

In this section, I defend a new, stronger version of the external condition of freedom. One does something freely, I argue, only if one does it for its own sake, that is, on account of its perceived intrinsic value. I show that this condition is consistent with the possibility of free instrumental action, and I examine what it would be for instrumental action not to be *merely* instrumental.

1.1

Many accounts of freedom as autonomy, as we have seen, begin with some version of the idea that acting freely means, at a minimum, doing what one wants to do. What is at stake here? What value or concern is this condition meant to articulate? I want to suggest that the relevant sense of what one *wants* to do should amount to more than a matter of what one *intends* to do. As we saw in the previous chapter, even one who acts under coercion can be doing what she intends to do. Yet, on the face of it, one wants to say that there is surely some sense in which an agent in this position precisely does *not* do what she "wants" to do—and that *this* is why she acts unfreely. If there is something to this impression, the kind of desire which is relevant for freedom must be more robust than that involved in mere intention.

This suggestion accords with the diagnostic observations I made at the end of Chapter II. Frankfurt suggests that satisfying the external and internal conditions of freedom corresponds to possessing "freedom of action" and "freedom of the will,"

respectively.[1] This seems quite plausible. Whereas freedom of the will, in the relevant sense, is a matter of the relation of the will to some standard of coherence, rationality, goodness, or whatever, freedom of action is a matter of the relation of action to the will. Of the externally coerced person who has not been brainwashed and who suffers no "inner defeat" but does precisely what she thinks on reflection she ought to do, given the unfortunate incentives she faces, one wants to say that it is not her will which is unfree but her action. But since this agent does what she intends to do, freedom of action must amount to more than doing what one intends to do.

The introduction of higher-order desires, we saw, does not provide the missing link so much as change the subject. But there is another, more promising distinction between different modes of desire (or of wanting or valuing, as I will sometimes put it): that between intrinsic and extrinsic desire. To desire something intrinsically is to desire it on account of its perceived intrinsic value (or goodness, as I will sometimes put it)—value which the object has in its own right and which does not reduce to the value of something *else*. To desire something extrinsically is to desire it on account of its perceived extrinsic value—that is, entirely on account of the perceived value of something else. To desire something both intrinsically and extrinsically, accordingly, is to desire it partly on account of the value of something else and partly on account of some value which the object itself has above and beyond the value of other things.

This distinction in desire maps onto the distinction between activity as a final end, done for its own sake, and instrumental action, done for the sake of some further

---

[1] Frankfurt, "Freedom of the Will and the Concept of a Person," 136.

end. To do something *purely* instrumentally—as a mere means to a further end—is to do it purely from an extrinsic desire, because the value which one finds in such an action, and on the basis of which one chooses it, reduces entirely to the perceived value of the further end which the action serves. For instance, when a sailor throws goods overboard in a storm purely in order to save the ship and his life, the action of throwing goods overboard has no perceived intrinsic value at all but is valuable only on account of the value of saving the ship and surviving. When one does something as a final end, on the other hand, one does it from an intrinsic desire to do it. To the extent that one's aim in acting is not merely something other than the activity itself, the value which one pursues in acting thus must inhere in the activity itself and cannot be reducible to something else.[2]

Once we have drawn these distinctions, we can see that it is the actualization of intrinsic desire which is really at stake in our pursuit of autonomy. We can appeal here to familiar considerations of identification and alienation. An autonomous action, it seems, ought to be an expression of its agent's will; the agent ought to be able to recognize in the action her own practical identity, those practical commitments which make her "who she is." But these commitments can only be her final ends, the ends which she pursues for their own sake and not *merely* for the sake of other things. The actions which we do merely instrumentally and value merely extrinsically express not our practical identity but merely the fact that they happen to be necessary for whatever

---

[2] Korsgaard argues that even action for its own sake is not always intrinsically motivated. Christine M. Korsgaard, "Two Distinctions in Goodness," *The Philosophical Review* 92, no. 2 (1983). I consider her argument below.

it is that we actually care for.

For example, someone who works in a warehouse or at a call center all day and is indifferent to their job, can be said to "want" to work there only as a means to further ends. It would be a cruel joke to say of someone in this position that their work expresses who they are as an agent and a valuer. Consider T, a call center specialist in the U.K. interviewed by Joanna Biggs a few years ago. T helps customers fix problems with their phone and broadband connections; essentially, he spends many of his waking hours processing routine complaints from strangers under strict supervision by management. There is a very striking moment in the interview:

> Would T be happy to think that his identity came from what he does all day? "I really hope not. I could not say enough how I hope not. I used to like who I was, and if this place is now my identity, then I don't like myself. Literally, apart from the few people that I can sit and have a chat with and a gas with, the money is only just passable as the reason I come here. So, if the money changed, or certain people didn't work here any more, I can safely say I would probably be at the Job Centre looking."[3]

We will meet T again later. What I want to highlight now is the immediate and natural connection which he draws between the fact that his work in some way fails to express his identity and the fact that it is a *mere* means to other ends. To be sure, he hedges a little, speaking of who he *hopes* he is rather than who he *is*. But this ambivalence simply reflects the fact that there is *some* sense in which "who he is" is a matter of his actual doings—what he "does all day." Evidently, the "self" constituted by these doings, though, is not one with which he identifies. And he explains why: it is not exactly that he *disavows* any of his own first-order desires; it is rather that the relevant desires are

---

[3] Joanna Biggs, *All Day Long: A Portrait of Britain at Work* (London: Profile Books, 2015), 87.

merely extrinsic and so fail to express his practical identity. We can thus do much of the philosophical work that Frankfurt and others do using the concept of higher-order desires, but while staying within the realm of first-order desire.

For what it is worth, my interpretation of our interest in autonomy is also consistent with empirical research on human motivation. For instance, consider the very robust finding that extrinsic rewards such as monetary incentives tend to undermine intrinsic motivation for activities that are "their own reward."[4] The best-supported explanation of this tendency is "cognitive evaluation theory," according to which extrinsic rewards tend to undermine intrinsic motivation when and to the extent that they are perceived to "thwart the need for autonomy."[5] As Edward Deci *et al.* point out, this theory invokes a certain understanding of "the real meaning of intrinsic motivation":

> Intrinsic motivation energizes and sustains activities through the spontaneous satisfactions inherent in effective volitional action. It is manifest in behaviors such as play, exploration, and challenge seeking that people often do for no external rewards. It is thus a prototypic instance of human freedom or autonomy in that people engage in such activity with a full sense of willingness and volition.[6]

I mention this also in order to illustrate how natural this way of talking is when we are not doing philosophy.

---

[4] Edward L. Deci, "Effects of Externally Mediated Rewards on Intrinsic Motivation," *Journal of Personality and Social Psychology* 18 (1971); Edward L. Deci, Richard Koestner, and Richard M. Ryan, "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivations," *Psychological Bulletin* 125, no. 6 (1999). The latter is a meta-analysis of 128 studies examining the empirical effects of extrinsic rewards on intrinsic motivation.
[5] "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivations," 628.
[6] Ibid., 658.

Two notes on terminology before I move on. First, as I have mentioned, I will use the verbs 'want', 'desire', and 'value' interchangeably. (I also use the terms 'goodness' and 'value' interchangeably, but I take it that this is fairly uncontroversial.) I want to emphasize the association of the first two terms with 'value', in particular, because this term makes it clear that desire in the relevant sense is a practical and not an affective concept.[7] To desire something is not necessarily to enjoy or take pleasure in every minute of it. An artist might want to work on her mural not in the sense that this activity gives her pleasure on a particular occasion—she might not feel like it on a given day—but in the sense that she values this activity intrinsically. (I will return to the topic of action for pleasure or enjoyment in Section 6.)

Second, I employ the locutions of 'perceived value' and 'perceived goodness' deliberately so as to make room for a philosophical account of desire as a response to objective value, rather than as the invention or construction of value. In fact, for ease of exposition, I will even tend to drop the qualifier 'perceived' on occasion. That said, for the rest of this dissertation, I proceed on the understanding that, though we are now dealing with freedom and unfreedom only in their external dimension, a full account of freedom would require in addition an account of its internal dimension. I will thus abstract from the fallibility of human desire purely for the sake of argument. The point is that when we find ourselves coerced rather than manipulated or brainwashed, and when we encounter external forces and obstacles rather than internal ones such as phobias or addictions, our unfreedom stems not from the fallibility of our desire but

---

[7] Thanks to Amichai Amit and Diane Perpich for pressing me on this point.

from its failure to find expression in our activity.

1.2

To do something instrumentally yet autonomously, one would have to do it not *merely* instrumentally but also for its own sake. One would have to do it on account of both its extrinsic value and its intrinsic value. In this section, I make an observation about the relation that has to obtain between these two values for activity motivated in this way to be possible.

When one performs instrumental activities such as cooking, bricklaying, or teaching, one characteristically does these with a view to distinctive further ends—producing (tasty) means of nourishment, constructing (aesthetically pleasing) shelter, and developing someone's capacities, respectively. But one can do each of these things simultaneously for its own sake—for instance, as the exercise of a valuable skill, or as the making of a valuable social contribution. If one does cook, lay bricks, or teach partly on account of some such intrinsic value, then one will feel that *something* of value would be lost if we had no need of food, shelter, or learning—if we were not vulnerable to the elements, say, or if we were born with all of our abilities fully developed. Or again, something of value would be lost if the relevant needs could be satisfied without labor—for instance, if our means of nourishment simply fell from the heavens in the form of mana (or Soylent), or if we learned everything automatically and without guidance.

What is it for an activity to have this sort of "double" value? For an activity to be intrinsically valuable, the intrinsic value must inhere in the activity in its own right; it cannot reduce to the value of something else. But this means it has to belong to the

activity by virtue of its being the activity it is; otherwise, it would belong at best to something else which merely happened to be attached to the activity in some way. In the case of instrumental activity, however, *what the activity is* is determined by its further end. If an intrinsic activity is good as such, then an intrinsically good instrumental activity has to be good as the instrumental activity it is. Its intrinsic value has to attach to it by virtue of its being useful in this distinctive way, that is, on account of its concrete instrumental value. This means that the intrinsic value of instrumental activities, though not *reducible* to the value of something else, can nonetheless be said to be *dependent*, *parasitic*, or *conditional* on the value of something else.

For example, one cannot understand what it is to cook except by reference to the end of producing means of nourishment (or something along those lines). Therefore, to value cooking intrinsically is to value it under this description: as the activity of producing means of nourishment. And in that case, the intrinsic value of cooking must depend on its instrumental value, since this instrumental value is the activity's usefulness for a certain end, which just is the aspect under which one values it intrinsically. For instance, one can intrinsically value cooking as the exercise of a valuable skill, but the relevant skill cannot be understood except in terms of its aiming at the production of a further good.[8] Similarly, one can intrinsically value cooking for

---

[8] This sort of idea crops up in the Marxian literature on "self-realization," where it is sometimes emphasized that the self-realization of the worker in her labor is essentially a "byproduct" of her aiming at other goods. See, e.g., Jon Elster, "Self-Realization in Work and Politics: The Marxist Conception of the Good Life," *Social Philosophy and Policy* 3, no. 2 (1986); and Arthur Ripstein, "Rationality and Alienation," *Canadian Journal of Philosophy* Supp. Vol. 15 (1989). I agree with Ripstein, *pace* Elster, that self-realization need not be *merely* a byproduct but can *also* be aimed at in its own right.

another person because this activity helps constitute a valuable form of social relation—but it could not help constitute this social relation if it were not a way of satisfying someone's need, that is, if it were not in the relevant way productive of a further good.

These observations will become directly relevant later on. In the meantime, though, they are also of some philosophical interest in their own right, for they reveal something important about the nature of intrinsic value, namely that such value is not necessarily context-independent and unconditional. In an influential paper, Korsgaard equates intrinsic value with unconditional value, which a thing possesses "if it is good under any and all conditions, if it is good no matter what the context."[9] She leverages this equation to argue that the distinction between valuing something intrinsically and valuing it extrinsically is orthogonal to the distinction between valuing something as a final end and valuing it instrumentally. Her point is that one can do something for its own sake and yet fail to do it on account of its intrinsic value, since one's final end might easily fail to satisfy the demanding requirement of being valuable in all contexts or unconditionally.

But Korsgaard's equation of intrinsic with unconditional value is untenable. Julie Tannenbaum has suggested numerous types of counter-examples. For instance, the presence of certain conditions can defeat the intrinsic value of a thing: "listening to Beethoven is intrinsically good, but nevertheless it is not good to do when someone is screaming for help outside one's door."[10] Likewise, certain conditions can enhance the intrinsic value of a thing without being the source of this value: "in the case of honors, it

---

[9] Korsgaard, "Two Distinctions in Goodness," 178.
[10] Julie Tannenbaum, "Categorizing Goods," *Oxford Studies in Metaethics* 5 (2010): 271.

is often the case that the more rare the honor is the greater the goodness of the honor. But rarity itself has no value."[11] My argument above suggests a different kind of counter-example. Sometimes, the intrinsic value of a thing is merely conditional because the thing itself—the bearer of value—depends for its identity on certain further conditions, and in particular on its relation to other things of value. The intrinsic value of cooking depends on the usefulness of cooking for further goods because cooking would not even be what it is in abstraction from this usefulness; yet the intrinsic value of cooking is not reducible without remainder to its usefulness.

2. THE UNFREEDOM OF MERELY INSTRUMENTAL ACTION

My account of external freedom promises to make sense of paradigm cases of external unfreedom such as action performed under interpersonal coercion or under duress by hostile circumstances such as natural disasters, illness, or public health emergencies. The account also brings into view further, distinctive kinds of external unfreedom which will be discussed at length in later chapters but which I introduce here.

2.1

The idea of freedom as activity for its own sake promises to provide a unified explanation of diverse classes of paradigm cases of unfreedom. In particular, it promises to explain both the unfreedom of action taken under duress by inhospitable circumstances and the unfreedom of action coerced by other people.

The first kind of unfreedom is exemplified in what we can call "necessary evils."

---

[11] Ibid.

These can include action taken under compulsion by overwhelming forces of nature: throwing goods overboard in a storm, abandoning one's house because of a flood. They can also include action necessitated by personal illness or injury, such as undergoing surgery. And they can include measures taken because of public health emergencies, such as doing an inordinate amount of washing and cleaning and self-isolating at home during a pandemic.

As I said in Chapter I, it is very natural to speak of genuine unfreedom in all of these cases, even if, by virtue of their non-interpersonal character, they are not the most central cases. And my account shows us why these actions are unfree. Interestingly, we can already find the explanation in Aristotle's original discussion of the example of throwing goods overboard in a storm.[12] He observes that such actions seem voluntary in the sense that "they are worthy of choice at the time when they are done."[13] They are not mere movements such as reflexes, nor are they like the "actions" one might take in an uncontrollable fit of terror. They are genuine, intentional actions, performed like any other ordinary action in pursuit of one's ends. We are reluctant to call them voluntary without qualification, however, because "no one would choose any such act in itself."[14] This is essentially the account I have given.

Not choosing one's action in itself (if this means not doing it for its own sake), I have argued, just is the general form of external unfreedom. The reason one acts

---

[12] Aristotle, *Nicomachean Ethics*, , 1110a4–19.
[13] Ibid., 1110a12.
[14] Ibid., 1110a19. Aristotle ultimately concludes that such actions are more like voluntary actions than involuntary actions—but this is not to say that they are voluntary without qualification.

unfreely when one throws goods overboard in a storm, or abandons one's home in a flood, is that one does this as a *mere* means to a further end, such as survival. The same account applies to actions like undergoing surgery. No one would choose to undergo surgery for its own sake; the process is inherently risky, frightening, sometimes painful, and frequently largely unconscious in any case. It is something one does solely for the sake of further ends, such as health and survival (or another's survival, in the case of transplant surgery). Measures taken under duress by public health emergencies clearly admit of the same treatment. When a pandemic forces people to incessantly clean themselves and wash their things and environments, confine themselves at home, and practice "social distancing" in public, people do not do these things for their own sake. People choose these actions rationally in the pursuit of their own ends, and for the most part without any interpersonal duress, but they do them merely instrumentally, not because they want to do them in the robust sense that matters from the point of view of autonomy.

Furthermore, the account I have proposed also promises to make sense of the unfreedom of coerced action. This is a more difficult topic and will take up a whole chapter in its own right, but we can already descry one possible explanation. When one acts under coercion by means of a threat, one does what one is doing in order to avoid the threatened consequence. In other words, coerced action is a species of instrumental action. But it is obvious that it must normally be a case of *merely* instrumental action. For if the coercee were already motivated to perform the action for its own sake—if the action were "its own reward"—it would be unnecessary to motivate him further by means of a threat. But in that case, coerced action is normally unfree, and its unfreedom

is a species of the unfreedom of merely instrumental action.

Now, in fact, much more will have to be said. In particular, this sketch does not yet indicate what *distinguishes* the unfreedom of coerced action from the unfreedom of generic necessary evils, other than its distinctive etiology. But that will have to wait until Chapter IV. My point, for now, is that we do have at least the beginnings of what we have been looking for: an explanation of the various species of external unfreedom and their unity.

2.2

Beyond deepening our understanding of familiar paradigm cases of unfreedom, my account also promises to extend our understanding of unfreedom beyond these cases. In particular, it brings into view manifestations of fundamentally the same kind of unfreedom which are perhaps less obvious, though they are right in front of our noses and on reflection rather troubling.

One apparent implication of my account concerns toil—labor which is so impoverished in its content that it loses any intrinsic value and becomes entirely unchoiceworthy except as a possible means to other ends. We have already had occasion to reflect on the nature of T's work at the call center. In the same vein, recall Phil, the spot-welder quoted in the Introduction, who pushes a button all day long. It seems to me that no one would do this sort of work for its own sake, just as no one would throw goods overboard or undergo surgery for its own sake. While it is useful as a means to further ends, there is nothing intrinsically valuable about pushing a button all day long. But if such work is necessarily done as a mere means to other ends, then it is necessarily done unfreely. A life spent largely in the performance of this kind of work,

then, is a thoroughly unfree one. And if it turns out that the condition in which some devote their lives to such toil is avoidable, then its continued existence on a massive scale would seem to be an indictment of the prevailing social order. The prevalence of avoidable toil is the mark of a society hostile to the freedom of the individual.

Another apparent implication of my account concerns paid labor—meaning labor which not only receives some kind of monetary compensation but which is actually *motivated* by the incentive of payment. I mentioned above that if a coerced agent were already motivated to perform the coerced action for its own sake—if the action were "its own reward"—then coercion would generally be unnecessary to get her to do it. The same reasoning holds of other kinds of incentivization. In particular, one might think that if people were generally motivated to do their work for its own sake, they would not need to be incentivized to perform it by the offer of payment. *Insofar* as payment functions as an incentive, then, one might suspect that paid work is also a species of merely instrumental action. Now, if one is lucky, the pay will not function as an essential incentive, for one will have a job which not only pays the bills but which one happens to find intrinsically valuable. But in any case, my account of freedom already suggests at least the following conclusion: if one does one's work for the money alone, one does it merely instrumentally and therefore unfreely.

Trivially, this will tend to be true of paid toil, since such labor cannot reasonably be done for its own sake. Certainly this is the situation of Phil and of T. Phil, as we saw in the Introduction, declines to get a "better" job because he knows the pay would be

worse.[15] The same is true of T:

> He came to the call centre through an agency as a "stop-gap and eleven and a half years later I'm still here. So I fell into a trap." He thought it would be a good job for a year or so, but "the money just keeps you here."[16]

But a job done merely for the money need not be *inherently* toilsome—that is, as far as its content is concerned. Recall Virginia, the Filipina woman quoted in the Introduction who works as a nanny for a family in Florida. One day, she hopes, she will have enough money to go home and live in a house of her own with her family—but she has not seen her own children in two decades. Now, Virginia's job of helping to care for children is not intrinsically unchoiceworthy as such. It is not like mere button-pushing. Nevertheless, it seems fair to say of Virginia that *she* does this work as a mere means to a further end, the end of one day living with her own family in the Philippines. It also seems fair to say that her work is less than fully voluntary. And my account puts us in a position to understand why.

Again, much more will need to be said—for example, about the difference between the mere instrumentality of paid labor and the mere instrumentality of coerced labor, as well as about the difference between both of these and the instrumentality of labor in general. I will elaborate and defend the implications of my argument for paid labor and toil in Chapters IV and V, respectively. For now, I have merely wanted to introduce the basic ideas and to make a case for their *prima facie* plausibility.

---

[15] Terkel, *Working*, 162.
[16] Biggs, *All Day Long*, 83. T earns £10.90 an hour, which is between $13 and $14 in 2020 U.S. dollars. Ibid., 86.

3. UNFREEDOM AND NECESSITY

In this section, I show that my proposed account of external freedom illuminates the intuitive connection between freedom and necessity. I then introduce an important refinement to the account, arguing that the necessity and unfreedom of merely instrumental action comes in degrees and showing that this explains a distinction of independent interest.

3.1

In Chapter I, I mentioned that in ordinary thought and language we associate free action with doing what one wants to do, and unfree action with doing what one is in some sense forced to do—with acting under the guise of necessity. We are now in a position to vindicate this language and explain precisely what it means. I have already indicated in what sense externally free action requires doing what one wants to do. Here, I will identify the sense in which external unfreedom is a form of subjection to necessity.

According Hobbes, the necessity which is opposed to freedom must be understood in the strictest possible sense. To count as doing something unfreely, it must be impossible, absolutely and without qualification, not to do it. For example:

> when a man throweth his goods into the sea for *fear* the ship should sink, he doth it nevertheless very willingly, and may refuse to do it if he will; it is, therefore, the action of one that was *free*: so a man sometimes pays his debt, only for fear of imprisonment, which, because nobody hindered him from detaining, was the action of a man at *liberty*.[17]

---

[17] Thomas Hobbes, *Leviathan* (Indianapolis: Hackett, 1994), Part II, Chapter XXI, 137, emphases his.

Though it has the virtue of being principled, this view seems flat-footed, to put it mildly, on two grounds. First, as I pointed out *ad nauseam* in Chapter II, when one person coerces another to do something by means of threats, we surely want to say that the coerced action is unfree, even though the coerced agent could, strictly speaking, "choose" to suffer the threatened sanction rather than perform the coerced action. Second, an action which is absolutely necessary will in many cases not so much fail to be *free* as fail to be an *action* at all. One feels, then, that Hobbes must be wrong.

The argument of the previous sections vindicates this feeling. The necessity which is opposed to external freedom is a species of instrumental necessity. It is a kind of necessity for good, which, as Aristotle puts it, marks "the conditions without which good cannot be or come to be, or without which we cannot get rid or be freed of evil."[18] The necessity associated with external unfreedom, then, is not absolute but relative to an end.

Now, it might be argued that even in this relative sense, the means to an end can be called "necessary" only on the assumption that the agent has no reasonable alternative means to the same end. G.A. Cohen defends a version of this proposal to argue, in effect, that an individual in the position of T (the call center worker) is not "forced" to work for his employer, even if his work is a means to the end of avoiding starvation, since he could achieve the same end by means of the reasonable alternative of founding a startup instead.[19]

---

[18] Aristotle, *Metaphysics*, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton: Princeton University Press, 1984), 1015a.
[19] G. A. Cohen, "The Structure of Proletarian Unfreedom," *Philosophy and Public Affairs* 12, no. 1 (1983): 7ff. This is of course a common argument in defense of capitalism.

But this is too quick—not just because it relies on a dubious empirical assumption but for conceptual reasons. First, even in the presence of reasonable alternatives, we can reasonably say that the means to an end is necessary for that end in a qualified sense: it is necessary, *other things equal*. That is, *some* means is necessary for the end without qualification, and the particular means that an agent does take is necessary for the end assuming that other things are equal, that is, assuming she does not go out of her way to take some other means instead.

Second, this qualified, *ceteris paribus* sense of necessity is the one which is relevant for unfreedom. An instrumental action is rarely if ever a necessary means without qualification, as there are almost always alternative means, often reasonable ones, available at a sufficiently specific level of description. For example, a slave who obeys her master's command to cook dinner in order to avoid a beating might cook a stew, but perhaps she might equally have avoided a beating by making soup instead. She thus has an alternative to her chosen means. Nor is there anything as such unreasonable about making soup—other than, in this case, the fact of its unfreedom, which is what is in question. (If unfreedom is *defined* in part as a lack of reasonable alternatives, then the reasonability of the alternatives obviously cannot depend on their being free, on pain of circularity.) Yet surely the slave cooks the soup unfreely. Therefore, necessity in the sense which characterizes unfreedom cannot be incompatible with the existence of reasonable alternatives. It is necessity, other things equal.

Cohen's anti-capitalist twist on it is that workers *as a group* are still forced to work for capitalists, and that individual workers might refuse to join the petty bourgeoisie out of solidarity with their fellow proletarians.

Of course, as I have argued, not *all* necessity of this kind is associated with unfreedom. The happy cook who cooks not as a *mere* means but also for its own sake acts under the guise of instrumental necessity, but not unfreely. The problem is not instrumental necessity as such but only *mere* instrumental necessity. When we say that acting unfreely is a matter of doing what one *has* to do, we have in mind specifically a sense of *having* to do something which gets its sense in part from its opposition to *wanting* to do something. This is what the qualifier 'mere' captures. In summary, then, my account gives us an explanation of the kind of necessity which distinctively characterizes externally unfree action in general. It is (1) mere (2) necessity, other things equal, (3) relative to an end.

3.2

There is a further problem. The necessity of merely instrumental action is relative to its end. But in that case, it can seem that if the end is not necessary, then the means cannot be necessary either. One only *has* to take the means *if* one *has* to pursue the end. But one does not always have to pursue the end—some ends are optional. And if the end which one is pursuing happens to be optional, then to do something merely for the sake of that end is surely not to do it unfreely, because one could simply choose not to do it, by giving up the end. Therefore, it cannot be the case that merely instrumental activity is *as such* unfree.

It is one thing to work in an undesirable job merely for the sake of securing access to food, shelter, healthcare, and other essential goods. This is typically the situation of low-wage workers, so my view naturally seems most plausible in relation to

these cases. But consider, on the other hand, a millionaire who already has enough wealth to be able to retire and live comfortably for the rest of his life, but who has a dream of taking part in a commercial space flight. Now this is what a philosopher would call an "expensive taste," and the millionaire, let us say, chooses to put in sixty hours a week in a soul-crushing and objectively pointless job in order to finance his dream. While dreams are nice to fulfill, however, this individual's end hardly counts as *necessary*. His work, too, though it is merely instrumental, accordingly seems optional—and therefore voluntary—in a way in which the low-wage worker's does not.[20]

This is the sort of thinking which informs Cohen's insistence that an agent, to be free, must have *reasonable* alternatives:

> When I am forced to do something I have no *reasonable* or *acceptable* alternative course. It need not be true that I have no alternative whatsoever. At least usually, when a person says, "I was forced to do it. I had no other choice," the second part of the statement is elliptical for something like "I had no other choice worth considering."[21]

Thus, an agent whose alternatives are working or starving is forced to work—she has no other choice, for all intents and purposes—because the alternative of starving is not reasonable. By contrast, the wealthy space-flight enthusiast has a reasonable alternative to working, namely giving up on his expensive dream.

How can the reasonability or unreasonability of one's alternatives make the difference between being free and being forced? It seems to me that in the background

---

[20] Relatedly, there is a political difference between the cases: the low-wage worker has a stronger claim to others' assistance than the millionaire. Compare Elizabeth Anderson, "Optional Freedoms," in *What's Wrong with a Free Lunch?*, ed. Joshua Cohen and Joel Rogers (Boston, MA: Beacon Press, 2001).
[21] Cohen, "The Structure of Proletarian Unfreedom," 4.

here lies a certain conception of the distinction between what we might call "need" and "mere desire," or between "necessities" and "luxuries." Certain things—such as survival and the means to survival—are objects of need, or necessities. Others, however—such as flying into space and the means to doing so—are objects of mere desire, or mere luxuries. And whereas the former are in some sense non-optional or "given" to us, the latter are optional, discretionary, or in some distinctive sense genuinely "chosen."

We can find this picture at work explicitly in contemporary liberal thinking about labor and free time, by Robert Goodin and others. These theorists conceive of autonomy as the capacity to form and act upon one's own principles, particularly about how to spend one's time. Strikingly, they argue that autonomy in this sense is opposed not just to the extreme condition of chattel slavery but also to what they call the "slavery" of paid labor (at least insofar as it is required to finance a minimal standard of living), unpaid household labor, and the immediate personal care required to meet one's own bodily requirements. And the reason we are supposed to be able to say that agents lack autonomy in all of these cases is that the activities in question are not "discretionary" but necessary, not up to us but "*dictated* by the 'necessities of life'."[22]

3.3

This is a tempting picture, but it rests on a fundamental misunderstanding and exaggeration of the difference between "need" and "mere desire," and consequently

---

[22] Robert E. Goodin et al., *Discretionary Time: A New Measure of Freedom* (Cambridge: Cambridge University Press, 1998), 34, emphasis mine.

also of the difference between toil by the poor and toil by the wealthy. The difference is not a categorical difference between freedom and unfreedom but a difference of degree within the category of unfreedom.

The problem is this. On the one hand, the view fails to recognize that the pursuit of necessities is also a product of genuine choice. After all, we can *in principle* choose to forgo the pursuit of necessities such as survival just like we can in principle forgo the pursuit of comparative luxuries like flying into space. On the other hand, it is equally misleading to suggest that there is some distinctive sense in which luxuries, as opposed to necessities, are objects of *mere* choice or discretion, that is, choice evincing the complete absence of necessity. For when our space-flight enthusiast acts in pursuit of his end of flying into space, he acts in pursuit of a perceived good just as when the low-wage workers acts to preserve her life. In *either* case, to give up the means would be, other things equal, to give up a perceived good. If it is true that one has no choice but to do what is necessary to preserve one's life, then in qualitatively the same sense it is also true that one has no choice but to do what is necessary for other goods.

That does not mean that there is no difference here; it just means that the difference has been taken for granted, and therefore remained unexplained and misunderstood. We can make sense of the genuine difference between strictly "necessary" and comparatively "optional" ends—the distinction between needs and desires, or necessities and luxuries—as a difference of degree. In general, we can say that the degree of necessity of an end is a function of its importance. And the degree of necessity of the means to an end similarly depends on the importance of the end. Of course, we do not generally operate with a detailed and comprehensive ranking of our

ends, but in the relevant cases, clear comparisons are easy enough to make. For example, it is more necessary to secure one's child's welfare than to complete one's stamp collection, and the activities performed for the sake of the former end are more necessary than the activities performed for the sake of the latter, because the former end is more important than the latter.

It might be objected that life, in the sense of mere survival, being compatible with a completely vegetative state, in fact has little or no intrinsic importance at all—so that this account does not explain the distinctive necessity of the most basic necessities (food, water, shelter, etc.) after all. Survival, however, is not just an end in itself but a means to further ends. When we do something for the sake of survival, we do not do it for the sake of survival alone but also for the sake of these further ends. And since survival is in fact a necessary condition not just of this or that particular activity but of our power of agency as such, these further ends normally include virtually all of our ends. Generally speaking, then, the instrumental activities necessary for survival will indeed be necessary for a large range of very important intrinsic goods, and accordingly necessary to a very high degree. Other so-called necessities typically have the same character of being instrumentally necessary for a large range of further intrinsic goods, by virtue of being conditions of our very power of agency.

Now if an instrumental activity is done freely—that is, also for its own sake—then the importance of the further good to which it is a means is not a *problem*; it is nothing other than the measure of the activity's instrumental value. In unfree activity, however, when the activity is a *mere* means, the importance of the good to which it is a means, while conferring instrumental value on the activity, also becomes the heaviness

of a chain, whose weight is the measure of the agent's unfreedom. Merely instrumental activity, in other words, is unfree in proportion to the importance of the good to which it is a mere means. Thus, the fact that a particular instrumental activity is a means to keeping one's child alive rather than maintaining one's stamp collection makes it more instrumentally valuable, but normally this is not a problem, inasmuch as its role in raising one's child also makes the activity intrinsically valuable. By contrast, paying a ransom to save one's child's life—which one does *merely* instrumentally—is especially hateful, and to be more precise, especially unfree, as compared with, say, paying a ransom to save one's stamp collection.

My account of freedom as activity for its own sake, then, in conjunction with the concept of the importance of an intrinsic end, can accommodate and indeed helpfully *explain* the distinction between strict "necessities" and comparative "luxuries," and the corresponding distinction between the instrumental activities done in the service of securing these goods. The distinction, however, turns out to be a matter of degree and accordingly more fine-grained than a simple dichotomy. Moreover, it turns out to be rather misleading to speak of "luxuries" in general as "optional" or "discretionary"—that is, as unnecessary—at all. Qualitatively speaking, "luxuries," the objects of so-called mere desire, are necessary in exactly the same sense as "necessities," the objects of need. Reflection reveals not that merely instrumental activity is sometimes strictly unnecessary and freely chosen, but that it is necessitated—and unfree—in proportion to the importance of its end, the importance of the good which hangs in the balance.

Of course, the difference of degree may be decisive for political purposes. In a world in which there are still workers who toil for access to basic necessities, their plight

should take priority and it would make no sense to focus instead on the unfreedom of the toiling millionaire.

4. TOO MUCH UNFREEDOM?

"The light dove," writes Kant, "cleaving the air in her free flight, and feeling its resistance, might imagine that its flight would be still easier in empty space."[23] In defending the idea of freedom as activity for its own sake, I might seem to be making the same sort of mistake as Kant's dove. So much of what we do is done merely instrumentally that we cannot reasonably regard all of it as unfree, unless we are prepared to wish away the fundamental conditions of human existence. Or so one might argue. In this section, I consider a version of this objection in order to show that my account is not obviously excessively ambitious. The discussion will allow me to refine my account in a couple of ways.

4.1

Rahel Jaeggi has proposed what *sounds* like a conception of freedom as activity for its own sake. In her book *Alienation*, she suggests that freedom requires not just "self-determination" but also what she calls "self-realization," which she says is a matter of living one's life not as a mere means but for its own sake.[24] Now, I do not regard these two ideas as distinct. As I have argued, "self-realization" in this sense just *is* a part of

---

[23] Immanuel Kant, *Critique of Pure Reason*, trans. Norman Kemp Smith (New York: Palgrave Macmillan, 1929), A5/B8.
[24] Jaeggi, *Alienation*, 207.

self-determination. More interestingly, though, when Jaeggi discusses the concept of self-realization more concretely, she waters it down substantially. "We can speak of self-realization," she says, "only when we do the things that make up our life for their own sake, or more precisely, when we orient our life as a whole around ends that we pursue for their own sake."[25] Clearly, these two formulations are not equivalent. The second, "more precise" one is weaker than the first.

Jaeggi backs off from the stronger version of the ideal because she is very impressed with the apparent unavoidability of merely instrumental action not just in our society but in any conceivable form of human life:

> It is never possible to completely avoid action grounded in means-end reasoning (action aimed at achieving external ends). It is likely that the attainment of every end is preceded by chains of actions in which not every individual link has intrinsic value. Even in playing the piano, something I do for its own sake, I must do finger exercises that I perform only because of their relation to the external end (external to the exercises themselves) of acquiring the technical skills to play a Beethoven sonata.[26]

If our ideal of freedom is to be attainable and worth caring about, then, it seems we must moderate our ambitions. As Wittgenstein would say: "Back to the rough ground!"[27]

In an attempt to articulate a more realistic ideal, Jaeggi suggests:

---

[25] Ibid.
[26] Ibid., 208.
[27] "The more narrowly we examine actual language, the sharper becomes the conflict between it and our requirement. (For the crystalline purity of logic was, of course, not a *result of investigation*: it was a requirement.) The conflict becomes intolerable; the requirement is now in danger of becoming empty.—We have got on to slippery ice where there is no friction and so in a certain sense the conditions are ideal, but also, just because of that, we are unable to walk. We want to walk: so we need *friction*. Back to the rough ground!" Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe, 3rd ed. (Malden, MA: Blackwell, 2001), §107.

> The possibility of self-realization […] is threatened precisely when one gets caught in a teleological circle, a situation in which one does one thing only for the sake of another without ever connecting them to a final end, that is, to an end where one can no longer ask the question for what purpose I am doing this? When this characterizes an entire life, the result is a fatal structure: "We then understand our action and our current life as a *means* to the end of another action and life, in which only there are we really ourselves." Hence, although instrumental action is always a part of the pursuit of life goals, it is meaningful only if it passes over into an action that leads to a goal that is not itself another means to an end but is pursued for its own sake.[28]

Jaeggi is surely right that a life consisting entirely of merely instrumental activities which never even aim at anything of intrinsic value represents an extreme failure of self-realization, the pointless waste of a human life in which desire is "empty and vain."[29] But living one's life for its own sake cannot be merely a matter of avoiding this kind of empty "teleological circle."

Recall the case of Virginia, who works two jobs in Florida so that one day she will be able to live with her family in the Philippines. Given that she has not seen her children for two decades, it seems no exaggeration to say that Virginia does not really live her life for its own sake but that her "current life" is "a means to the end of another action and life," in which other life she would be "really herself." The problem, however, is not that she is stuck in an empty "teleological circle." She is not unable to answer the question, "For what purpose am I doing this?" On the contrary, fundamental features of her life, such as her work and where she lives, are organized quite consciously—perhaps more clearly than most people's lives—around an

---

[28] Jaeggi, *Alienation*, 208, quoting Friedrich Kambartel, "Universalität als Lebensform," in *Philosophie der humanen Welt* (Frankfurt: Suhrkamp, 1989), 24.
[29] Aristotle, *Nicomachean Ethics*, 1094a18–22.

intrinsically desired end, namely the end of living in a house with her family. Her life satisfies Jaeggi's weak condition of self-realization almost to the point of caricature. The problem is that this condition is insufficient for true self-realization, which requires not merely that the activities which make up one's life be done for the purpose of intrinsically desired further ends but that the very activities which make up one's life also be ends in themselves.

In a way, then, Jaeggi overreacts to what she takes to be a hopelessly demanding conception of self-realization by offering in its place a hollowed-out alternative which it would be too difficult to *fail* to satisfy. But of course this does not invalidate her worry about the more ambitious conception. What, if anything, can be said in its defense?

4.2

To begin with, we should draw one more distinction. For the purpose of ethical and political thought in general, we need to differentiate, to put it crudely, between actions that really matter—actions that in some way play a significant role in our lives—and actions that are comparatively trivial. This distinction is relevant when thinking about freedom in particular. In general, the unfreedom of independently significant actions is bound to be troubling in a way in which the unfreedom of completely trivial actions is not. As Charles Taylor puts it:

> we make discriminations between obstacles as representing more or less serious infringements of freedom. And we do this, because we deploy the concept against a background understanding that certain goals and activities are more significant than others.[30]

---

[30] Taylor, "What's Wrong with Negative Liberty," 217–218.

This will be so on any plausible account of freedom.

The fact that we effortlessly and uncontroversially draw such distinctions in the course of political philosophy emerges most obviously in the identification of certain *domains* of human activity as particularly important subjects of freedom or unfreedom: labor, sexuality, religion, political participation, and so on. Taylor, to illustrate his point, contrasts the restriction of individual freedom imposed by laws instituting traffic lights with the restriction imposed by laws curtailing liberty of religion. We are likely to object more strenuously to the latter, he suggests, because it intrudes in an independently more significant sphere of action.[31]

In addition to the domain of a given activity, though, the way in which we choose to describe it also matters. In studying the effects of a law curtailing freedom of speech, one might choose to focus on the ways in which the range of permissible motions of the larynx has been narrowed—but this description obviously misses the point; it fails to reveal the significance of speech and its freedom in social life. Good political philosophy is in part a matter of asking the right questions. When it comes to freedom, this means inquiring into the right domains of activity, but it also means studying activity under the right sort of description.

Now, it seems to me that the right sort of description of the pianist's finger exercises—one which reveals the role of this activity in the agent's life—would be something on the order of "playing the piano," something which the pianist *ex hypothesi* does for its own sake and therefore freely. The more minutely we describe the activity,

---

[31] Ibid., 218.

the more likely we are to lose sight of its intrinsic value—but at the same time, we become focused on an increasingly trivial subject-matter. Even *if* we were to find unavoidable unfreedom at some sufficiently minute level of description, this unfreedom would be insignificant for *independent* reasons, because the activity to which it attaches is, described in that way, trivial. For this reason, the insignificance of unfreedom at this level would appear to be a consequence of *any* reasonable ideal of freedom. It is thus not a reason to reject the ideal of freedom which I am defending.

Importantly, there are some contexts in which highly specific levels of description are appropriate because they do pick out actions of significance in the agent's life. For the spot-welder Phil, who pushes a button 12,000 times a day, minute finger movements are not merely a trivial part of his life but more or less the entirety of what he does all day. In this case, we cannot say that any unfreedom we may find at this level of description is bound to be insignificant for independent reasons. But I would not want to say so. My account suggests that Phil's button-pushing is unfree, and the fact that it occupies a large portion of his waking hours, and that his life is in large part organized around it, does make this unfreedom especially troubling.[32]

This response to Jaeggi's objection grants one of her key premises for the sake of argument. What really matters, certainly for political purposes, is freedom and unfreedom in non-trivial domains of activity and at non-trivial levels of description. But in that case, even *if* we agree with Jaeggi's suspicion that mere instrumentality is

---

[32] Phil's button-pushing is also independently significant for qualitative rather than quantitative reasons, inasmuch as it is what he does for *work*—an important category of activity in our social lives and identities.

unavoidable in human action, it does not follow that mere instrumentality is unavoidable where it really matters. The latter, stronger claim is not self-evident and would have to be demonstrated specifically.

4·3

Perhaps we could just leave it there. But in fact I think it is important to challenge Jaeggi's ethical presuppositions about what human action *must* be like.

Let us begin with her example of finger exercises. Are these really a case of merely instrumental action? Not necessarily. First of all, Jaeggi seems to suggest that the finger exercises are somehow a "preliminary" to the real activity of playing the piano, whereas in fact they are a part of this activity. It is true that they also aim at something beyond them, since they are a means to other parts of playing the piano. But this is insufficient to show that they are *merely* instrumental.

It may also be true that they do not tend to be the most satisfying or consistently enjoyable part of this activity in itself. But this also does not make them merely instrumental. For a start, it does not mean that they are never satisfying or enjoyable in their own way. There can be a distinctive enjoyment in the process of noticeably improving one's skills, and also in the exercise of skillful, attentive, precise motor control. In any case, even when one does not happen to take *pleasure* in the activity on a given occasion, one might still intrinsically value it under these aspects and do it partly on account of this intrinsic value. It therefore strikes me as quite plausible that someone who genuinely valued playing the piano might, in particular, play finger exercises partly for their own sake.

One way we can see this more clearly is by contrasting the finger exercises with

superficially similar actions like Phil's button-pushing. Both actions involve certain repetitive motions of the fingers, and one might even say that, as far as pure bodily motion is concerned, this is all they boil down to. Yet one wants to say that the pianist's finger exercises are clearly valuable in themselves in a way in which Phil's button-pushing clearly is not. They are, for instance, skillful in a way in which the button-pushing is not.

Why can this difference be so hard to see? I think this has to do with the conditionality or context-dependence of the intrinsic value of an activity like piano finger exercises. If we attend to these exercises in abstraction from the broader pattern of activity in which they are embedded, then we cannot see what is valuable about them. Indeed, without this context, they would not *have* that value. Similarly, the impossibility of doing an action such as Phil's button-pushing for its own sake seems to come clearly into view only when we attend to the broader context of any given button-pushing: its repetitiveness, for instance, and the lack of scope for deliberation on the part of the agent.

But as I argued earlier, we should not confuse intrinsic value with unconditional or context-independent value. A thing's value can be dependent on its standing in a certain relation to other things without thereby being entirely reducible to the value of other things. As I mentioned, one way this can be the case is if the thing itself depends for its identity on the context. That is just what we find here. The finger exercises would not *be* the action they are—quite different from Phil's button-pushing—if it were not for the broader activity in the context of which we can see their intrinsic value.

It might be objected that it is a special feature of Jaeggi's example that it involves

the development and exercise of skill. What about more mundane actions such as walking from one point to another during the course of one's day? It is not as obvious what the intrinsic value of such an action could be. Yet, no matter how awesome our powers may become, the occasional action of this sort is unavoidable in any conceivable form of human life, given that we are embodied creatures.

The notion of embodiment, though, itself suggests the key to one possible answer. For it seems to me that even actions such as walking from one point to another can be intrinsically valuable as exercises of our physical power of locomotion. This may be less impressive than the skill involved in piano-playing, but we might think, for instance, that one loses something of intrinsic value if one never moves around on foot but always has one's needs met wherever one happens to be, or gets everywhere by car or Segway. Consider dystopias depicted in films such as *Wall-E*, where future human beings, constantly and closely attended by robots, have lapsed into a state of virtually total passivity. It is not an inspiring vision.

Of course, not every life without walking is a life of passivity, even in respect of locomotion. And while I am suggesting that walking can have a distinctive intrinsic value, which one does not actualize if one does not or cannot walk, it does not follow that it is the only intrinsically valuable way of getting around, or that it is more valuable than the alternatives. In particular, it does not follow that a disability which prevents one from walking makes one worse off. A disabled person who cannot walk can be expected, at least in favorable social conditions, to have access to a slightly different set of intrinsic goods, some of which are likely in turn to be unavailable to non-disabled people. The disability rights activist Harriet McBryde Johnson speaks of how "it's a

great sensual pleasure to zoom by power chair on these delicious muggy streets."[33] This is consistent with and in fact reinforces the point I am making, which is that even mundane and seemingly uninteresting activities, such as a certain way of mobilizing one's body in order to get from one place to another, can be valued intrinsically.

On the other hand, this does not mean that all of our activities are intrinsically valuable just because they are embodied—Phil's mindless and repetitive labor being a case in point. No doubt more needs to be said about what exactly makes such labor distinctively undesirable. More generally, a full account of the implications of my conception of freedom is bound to depend to some extent on a determinate system of substantive ethical values. I have made no attempt to fully articulate, let alone defend, the various values to which I appealed in this discussion. I have merely indicated the sorts of values to which one might quite reasonably try to appeal. I will return to the case of routine labor, in particular, in Chapter V.

For now, my point is just this. When Jaeggi asserts the unavoidability of merely instrumental action, she is not stating a self-evident metaphysical truth grounded in nothing but the form of human action as such. She is rather making a substantive and debatable ethical claim, which depends on some determinate system of values no less than its negation does. To see whether her claim is true, much more work will have to be done. For now, therefore, I will proceed on the presumption that my view of freedom is not unreasonably demanding.

---

[33] Harriet McBryde Johnson, "Unspeakable Conversations," *New York Times Magazine*, February 16 2003, quoted in Elizabeth Barnes, *The Minority Body: A Theory of Disability* (Oxford: Oxford University Press, 2016), 137. See also ibid., 56–58.

5. PLEASURE AND PRINCIPLE

I have suggested on several occasions that doing something for pleasure is not the only way of doing it for its own sake—but is it at least *a* way of doing it for its own sake? And what about action done from a sense of moral duty, which is sometimes taken to be a paradigmatically autonomous exercise of the will?

5.1

There are two ways of understanding what it is to do something for pleasure or enjoyment, and just one of these corresponds to a way of doing it for its own sake.

On one tempting picture, pleasure is essentially an affective state independently intelligible of that which causes it, or of its object. If this is all pleasure comes to, then to take pleasure in some activity is to be caused to be in this affective state when one engages in the activity. The state, however, is entirely distinct from the activity and could in principle be caused in some other way: by doing something else, perhaps, or simply by taking a drug or being hypnotized.

To do something solely for the sake of pleasure in this sense is to do it merely instrumentally. One's end is not the activity itself (or any activity at all) but *being in a state of pleasure*, and it just so happens that one's activity is a means to that end. Indeed, the activity may not even coincide with the end but may precede it. And if—as is in principle possible—some other activity were found to produce the same end, it would not matter in the slightest which activity one chose (other things equal). The value which one is after is entirely reducible to the value of the further end, and the means itself is a matter of indifference.

Now, it does seem obvious that there are such things as affective states of this kind—brute feelings whose identity is independent of their object or cause—and it does seem that we can act for the sake of coming to be in such states. I suspect that action done for the sake of producing brute physiological states with certain affective components are the likeliest candidates. And it is strictly true that when one acts in this way, one does what one is doing under the guise of mere necessity and hence unfreely. In the context of a reasonably good human life, such desires (e.g., the desire for one particular kind of physiological sensation) and the actions taken in their pursuit (e.g., occasionally swallowing a pill) can be expected to play only an insignificant role, so these actions are unlikely to be cases of anything but utterly trivial unfreedom. In principle, though, they can become sites of significant and troubling unfreedom, as when someone is constantly in pain and depends on the physiological production of specific affects (by simply taking a pill or perhaps by means of a laborious process of physiotherapy) to put them in a halfway tolerable condition.

In most lives, though, I suspect that the pursuit of pleasure of this *purely* affective type is at most a very niche side interest. Normally when we do something for pleasure, we do not do it merely in order to, as it were, "hack" our brains so as to produce some independent mental state. Normally, we take pleasure *in the activity itself*. That is, normally when we do something because we enjoy it, what we are after is not some separable result of the activity alone but rather the (pleasurable) engagement in that very activity. We could call this "practical pleasure." For instance, someone who *enjoys* going swimming and swims for that reason does not aim to produce some state which could in principle equally be produced by doing something else instead; what she

wants is precisely to swim. But this looks like a case of activity for its own sake.

It seems to me that many cases that *look* like the pursuit of pleasure as a pure affective state are actually at least partly the pursuit of genuinely practical pleasure. Consider, for example, the recreational consumption of drugs. Since this activity is likely to have at least some sort of social, cultural, or even private ritualistic element, it is likely that what one is after is not *merely* a pure affective state but also to some extent the activity itself. If a philosopher came along and offered one a pill that could give one a sensation *as of* taking drugs with one's friends—it would not be the same.

5.2

What about action from principle—in particular, action done from a sense of moral duty? According to a venerable tradition, morally motivated action is an exemplar of autonomous activity. On my view, this can be true only in a partial sense.

Kant's argument in the *Groundwork* is that the will, understood as practical reason, is free if and only if it determines itself according to its own laws rather than being determined by alien natural causes in the form of mere "inclinations," and that this condition is satisfied if and only if one acts from respect for the moral law.[34] Thus, the agent who acts from a moral motive acts freely—and this is true even if her action is a "necessary evil" and chosen only as a means to further ends. In his *Lectures on Ethics*, Kant argues along similar lines:

> no man willingly gives away his possessions, but if he can save his children no otherwise than by the loss of what he has, then he does it, and is here practically necessitated. Hence a person constrained by motivating

[34] Kant, *Groundwork of the Metaphysics of Morals*, 4:446–447.

> grounds of reason is constrained without it conflicting with freedom. We perform these actions reluctantly, indeed, but we do them nevertheless, because they are good.[35]

This view, however, is obviously false if it is understood as a completely general conception of freedom. For it is possible for an action to be both coerced and motivated by a moral principle. But a coerced action is necessarily unfree. Therefore, moral motivation cannot be sufficient for freedom. One *might* pay a ransom to a kidnapper in order to save his victim on account of some non-moral motive—perhaps one simply does not want to tarnish one's reputation. But it is obviously possible to pay the ransom from a moral motive—say, in order to fulfill a duty of assistance to the victim—and this action would obviously still be coerced and therefore unfree.

This does not mean we have to reject the *Groundwork* view altogether. In that work, Kant is concerned with the internal conditions of freedom. We can (if we want to) accept his claim that acting from respect for the moral law is a necessary or even a sufficient condition of internal freedom without supposing that this condition tells us anything at all about external freedom. Even *if* moral motivation plays a special role in the internal dimension of freedom, it is not a power which can magically destroy external unfreedom.

I may seem here to be picking a fight not just with Kant but also with an older tradition of thought about ethics. According to Aristotle, virtuous actions must be

---

[35] *Lectures on Ethics*, trans. Peter Heath (Cambridge: Cambridge University Press, 1997), 27:268. Gerald Dworkin uses essentially the same argument to reject the sort of view I am defending. See "Acting Freely," 377.

chosen "for themselves."[36] The example above, however, also shows that this sort of principle cannot be right if it is taken to mean literally that any action which is virtuous thereby counts as being done for its own sake. When one pays a ransom to save a kidnapping victim, no matter whether the desire to save the kidnapping victim springs from genuine virtue or from some other disposition or desire, it is surely absurd to say that one pays the ransom for its own sake. This action has nothing to recommend it in itself, as the action it is; it is good purely on account of the further end that it brings about.

Aristotle's claim comes in the context of a discussion of the difference between doing what the virtuous person would do and doing it as the virtuous person would do it. What he has in mind—and what makes his suggestion plausible—is surely the weaker claim that if a genuinely virtuous action has to be done on account of its virtuousness and not on the basis of an ulterior motive. For instance, if paying the ransom is the right thing to do on account of its being the just thing to do, then virtuously paying the ransom means paying the ransom because that is the just thing to do: one chooses the just action "for itself" in the sense that one chooses it *qua* just, and on the basis of the considerations that make it just.[37] The pursuit of the end of *justice*, presumably, cannot be merely instrumental. Justice must be valued for its own sake and

---

[36] Aristotle, *Nicomachean Ethics*, 1105a32. Compare Candace Vogler's interpretation of Aquinas in her *Reasonably Vicious* (Cambridge, MA: Harvard University Press, 2002).
[37] See, e.g., G. E. M. Anscombe, "Good and Bad Human Action," in *Human Life, Action and Ethics: Essays by G. E. M. Anscombe*, ed. Mary Geach and Luke Gormally (Exeter: Imprint Academic, 2005), 198–199; Jennifer Whiting, "Eudaimonia, External Results, and Choosing Virtuous Actions for Themselves," *Philosophy and Phenomenological Research* 65, no. 2 (2002).

not, say, because behaving justly is the best way to appear to be just and gain a reputation for justice. But it does not follow that those actions which are done for the sake of justice are themselves done for their own sake.

It may be true that in a good world, the actions required by virtue or morality would indeed generally be done for their own sake. In a good world, one would assist one's fellows, for instance, not by paying ransoms on their behalf but by satisfying their needs through instrumental activities that one could simultaneously do for their own sake. It is only in this world that we can expect morality or virtue to be consistent with freedom, and one could say that it is really this world which our agency is "meant for."[38] But in many ways it is not now the world in which we live. In the actual world, living a moral or virtuous life is not always compatible with freedom.

---

[38] It would be the social and material condition of something similar to what Kant calls the "highest good" for a person. Immanuel Kant, *Critique of Practical Reason*, in *Practical Philosophy*, ed. and trans. Mary Gregor (Cambridge: Cambridge University Press, 1996), 110.

CHAPTER IV. COERCION AND EXPLOITATION

In this chapter, I give an account of the interpersonal dimension of external unfreedom, beginning with coercion by means of threats. A coercive threat, I argue, essentially provides its recipient with an incentive which is radically extrinsic to the action which it incentivizes (Sections 1–2). For this reason, coerced action is necessarily *merely* instrumental (Section 3). Given my account of freedom from the previous chapter, then, I explain the unfreedom of coerced action as a species of the unfreedom of merely instrumental action in general, and I explain the *pro tanto* wrongfulness of coercion on this basis; in addition, having brought coerced action under a broader genus, I identify its specific *differentia* and thereby explain what is distinctive about coercion vis-à-vis other sources of unfreedom (Sections 4–6).

A certain kind of exchange involves the same radically extrinsic incentivization as coercion; it follows that this kind of exchange, though commonly considered morally innocent, is in fact also unfree and *pro tanto* wrongful. I defend this implication, bringing both coercion and the relevant kind of exchange under the heading of "using someone as a mere means," "unfairly taking advantage of someone's vulnerability," or "exploitation" (Sections 7–8). The moral difference between coercion and the relevant kind of exchange, I argue, is a matter of the scope of the exploited agent's vulnerability (Section 9).

1. COERCIVE INCENTIVES AS RADICALLY EXTRINSIC

Coerced action, being performed for the sake of some coercive incentive such as the avoidance of punishment, is a species of instrumental action, performed not purely for its own sake but for the sake of a further end. It is true of all instrumental action that its further end is strictly distinct from, and in this sense external to, the action itself. In the case of coerced action, however, the end is external to the action in a distinctively radical sense. As I will put it, the end is a *radically extrinsic* further end.

1.1

In general, we understand what a given instrumental action is—what it is to perform that action rather than some other one—by reference to its characteristic further end. The association with this end is what makes the action the action it is. We can thus identify certain actions as characteristically aiming at certain ends just by virtue of the relevant action concept. In this way, we can say, for instance, that cooking characteristically aims at nourishment, medicine at health, housebuilding at the provision of shelter, and lecturing at the imparting of knowledge. The thought of the *what* already contains, and indeed depends on, some thought of the *why*.

Let us see how this works in more detail. What is it, say, to cook a meal for someone? Characteristically, it is intentionally to produce a means of nourishment for that person. There are aesthetic and affective ends, too, such as the enjoyment of good flavor, but for the sake of illustration, let us first consider nourishment alone and complicate the picture later. The production of the good of nourishment, then, is what cooking aims at—what it is for. It gives the action its point and specifies its success-

conditions. But it is also what gives the action its *identity*. That is, the characteristic end is what distinguishes *cooking a meal for someone* from other actions, such as *concocting a perfume*, *sterilizing tools*, *sending smoke signals*, or *destroying evidence*. Characteristically, however, intentionally to produce a means of nourishment is not merely to produce something which, as it happens, enables or contributes to the end of nourishing someone. It is specifically to produce something *with a view to* that end. One cooks *for the sake of* enabling or contributing to someone's nourishment.

We can go further. Nourishment itself essentially being a means to the maintenance of the body, to aim at enabling or contributing to someone's nourishment is to aim at enabling or contributing to the maintenance of that person's body. Bodily maintenance, in turn, is a means to the various ends which require the use of the body. The requirements imposed by the pursuit of these further ends determine what counts as the successful maintenance of the body: an Olympic sprinter will require a different diet than a Sumo wrestler, or a chartered accountant. We can conclude, then, that in cooking a meal for someone, one characteristically aims ultimately to enable or contribute to that person's pursuit of their ends.

To say that a given action is *characteristically* done with a view to certain ends is to say that this is what it is to perform that action without qualification. The characteristic ends give the action its identity; without them, the action concept would lack content. In this sense, even though they may be further ends strictly distinct from the action, they can nonetheless be said to "belong" to the action, or to be in *a* sense "intrinsic" to it. If one were explaining the action—what it is to do that thing—to an alien, one would have to refer to the characteristic end or series of ends. And if one is

told simply that a given agent performed the action in question on some occasion, one is generally entitled to assume, absent evidence to the contrary, that the agent performed it for the sake of the characteristic end or series of ends.

1.2

This does not mean, however, that things can never be otherwise in particular cases. Consider how cooking a meal for someone *under coercion* differs from the standard case described so far. As we have seen, one who cooks a meal under coercion still does so intentionally, because cooking a meal, under the circumstances, promotes some end of hers. But her end is not the nourishment of her coercer, or the enabling of the coercer's pursuit of his own ends. It is rather the avoidance of the consequence—such as a beating—which the coercer threatens to impose upon her.

But an end like avoiding a beating is not internally connected with the action of cooking a meal in the way that the end of enabling someone's nourishment is. Avoiding a beating is not what cooking is for; it is not the good which cooking *as such* produces. Cooking, as we have seen, is good for providing nourishment. *Learning self-defense* is good for avoiding a beating. But cooking is not good for avoiding a beating. The coerced agent, then, though she deploys the same action concept as the uncoerced agent, does not deploy that concept in the standard, unqualified way. For she does not herself aim at the characteristic end of her action—the end which belongs to the action by virtue of its being the action that it is. Her further end, we can say is *radically extrinsic* to her action.

Reflection on the nature of threats shows that this must be true of coerced action in general, at least where such coercion operates not by manipulating the agent's inner

desires but through straightforward external threats. For if a given end is intrinsic to a given action, then it is one of the goods which the action can be expected to produce anyway (if all goes well), just by virtue of being the action that it is. But merely to inform someone that her action will anyway contribute to that end, or to warn her that her refusal would undermine the end, is not to threaten her. Even if one informs or warns her of this fact with a view to influencing her behavior, this is not a case of coercion but merely a case of persuasion.

For instance, suppose I am in need of nourishment but unable to provide it for myself, and that no one else, other than you, is willing to provide it for me. Suppose I inform you of my predicament, making it clear that my need for nourishment will be satisfied if and only if you cook for me. Other things equal, telling you this does not count as the issuing of a *threat*, even if I do it *in order to* convince you to cook for me. By the same token, other things equal, if you do choose to cook for me on this basis, you do not do so under coercion. Even if I had been prepared to coerce you, coercion in this case would have been unnecessary since you are sufficiently motivated to perform the relevant action on the basis of the good that it brings about anyway.

1.3

As the public imposition of a conditional extrinsic good or bad for the purpose of influencing another's will, we can say that a threat is a species of "incentive." There are other types of incentives, including certain types of offers, as when one agent, *A*, offers to produce (or otherwise provide) good *a* for *B* so as to induce *B* to produce (or provide) good *b* for *A* in return. Here, too, the incentive must be extrinsic to the incentivized action. Given that the prospect of obtaining *a* is sufficient to motivate *B* to produce *b* for

A: if obtaining *a* were an *intrinsic* result of producing *b* for *A*, then *B* would *already* be motivated to produce *b* for *A*, and would not need to be further incentivized to do it. In these circumstances, if *A* were to induce *B* to produce *b* for *A* by informing *B* that producing *b* for *A* would assure *B* of obtaining *a*, this would not be the making of an offer or the issuing of any kind of incentive but (other things equal) merely an act of rational persuasion.

In the literature on the use of incentives in public policy, the meaning of 'incentive' is sometimes restricted to offers only (and not threats). I will not use the term in this narrow way. But putting aside this superficial terminological question, it is worth noting that this literature regularly invokes what amounts to the conceptual distinction between characteristic and radically extrinsic ends which I have been describing. Ruth Grant's definition of incentives, for instance, includes a necessary condition to the effect that an incentive is an "*extrinsic* benefit" that is not, as she puts it, "the *natural or automatic* consequence of an action."[1] And she points out, much as I have done, that

> [i]f the desired action would result naturally or automatically, no incentive would be necessary. An incentive is the added element without which the desired action probably would not occur. Thus, it is central to the core meaning of incentives that they are an instrument of government in the most general sense: they aim to direct people's behavior.[2]

---

[1] Ruth W. Grant, *Strings Attached: Untangling the Ethics of Incentives* (Princeton: Princeton University Press, 2012), 43, emphases mine.

[2] Ibid., 43–44. We might also refer back once more to the literature on extrinsic rewards and intrinsic motivation, and particularly the distinction drawn in some studies between "exogenous" rewards, "which are not an inherent part of the task" and which have been found to undermine extrinsic motivation, and "endogenous" rewards, "which are inherent in the task" and actually increase intrinsic motivation. Deci, Koestner, and Ryan, "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivations," 643.

2. SOME COMPLICATIONS

In this section, I will briefly examine some complications for the distinction between characteristic and radically extrinsic ends. My hope is that this examination will bring into clearer view some of the peculiar features of radically extrinsic motivation, particularly their paradigmatically interpersonal character.

2.1

The characteristic ends of activities are usually plural and subject to contestation and interpretation. This complicates but does not undermine the distinction I have drawn.

As I mentioned, cooking does not aim *only* at nourishment. Cooking is for eating, but we eat not just under the guise of the nutritious but also under the guise of the delicious. Part of one's aim in cooking, then, is characteristically to produce a meal which tastes good. Indeed, we may be after a meal which not only tastes *good* but also reveals interesting and novel (or "challenging") flavors and combinations of flavors to us. This is an important source of creativity in cooking. More generally, cooking is not *only* for eating. Another aspect of a meal which we characteristically care about is its visual appearance. This will be more so in some cases than others—think desserts and wedding cakes.

A special source of pluralism is the way activities can often be done in derivative ways that transform their ends in some way or introduce other ends into the picture. For instance, one can cook as part of a process of teaching or learning how to cook, and one can cook as a demonstration of one's skill, such as in a competitive bake-off. In these special contexts, some of the characteristic ends of cooking—particularly the end of nourishment—may recede into the background, in the sense that the eating of the

particular product one creates is not really the point of its creation. In the process of teaching someone how to cook, or learning the skill by practicing it, one may in fact cook something which is edible and which one then goes on to eat. But in a way, this is not the point of what one is doing: the point is the development of the skill of producing such things, in oneself or in another.

However, this complication also does not undermine the very idea of a characteristic end. Activities such as teaching to cook are derivative parts of a more general craft or practice of cooking. They are derivative in the sense that their characteristic ends and success-conditions are essentially parasitic on the ends of cooking *simpliciter*. If one is cooking as a way of demonstrating the activity for the purpose of teaching, the point of one's activity may not be the eating of what one is making. But if what one makes is not the sort of thing that *would* be a good means of nourishment, one is not doing it well. Or perhaps one is not cooking at all but merely giving the appearance of cooking, which may be sufficient for some kinds of demonstration. But this likewise does not undermine the idea that *cooking for real* has certain characteristic ends.

A final complication under the heading of pluralism is that the characteristic ends of our activities are generally open to contestation and interpretation. It may seem that the very idea of a characteristic end amounts to a narrow and conservative essentialism. However, while one could say that there is a form of essentialism involved, there is nothing narrow or conservative about it (nor anything metaphysically occult, since we are talking about action, which is shot through with intentionality).

Consider a familiar debate concerning the "meaning" of sex. Some conservatives think that the only purpose of sex is reproduction, and that sex purely for pleasure is

therefore a defective form of the activity. Others roundly criticize them for their gratuitously narrow conception of the characteristic ends of this activity, and argue that sex can also, and alternatively, be for pleasure. Clearly, as a debate about the "purpose" of sex, this is not a debate about the further ends for the sake of which it is *possible* to have sex. The conservatives accept (reluctantly) that people do have sex for pleasure. The debate concerns the "purpose" of sex—what sex is "for"—in the sense of what end or ends give the activity its identity conditions. Those who argue that sex is for pleasure thus make an essentialist claim just as their opponents do, though their essentialism is more pluralistic.

Now, perhaps one can put aside the "metaphysics" of action here and take an entirely deflationary view of this particular debate. Perhaps the conservatives and the liberals simply disagree about the moral significance of various kinds of sex and that is all there is to it. Be that as it may, it would be a definite mistake to claim, more strongly, that sex has no characteristic purposes at all. One of its purposes may be pleasure. More generally, its purposes may be plural and contestable; they may be subject to reinterpretation over time; they may vary from society to society; but none of this shows that *anything goes* when it comes to the question what sex is for. And this is because the good at which an activity aims is part of what gives that activity its identity.

We need *some* not entirely indeterminate sense of the characteristic ends of an activity in order to form a contentful concept of the activity at all. Consider the following. One can have sex for money. One can also cook for money. What makes doing the one different from doing the other? It is not the actual purpose for which the vendor does them: both are in fact done for money. What distinguishes them is primarily that they have different relevant *characteristic* ends. Cooking is for

nourishment (etc.). Having sex in this case is presumably for pleasure.[3]

2.2

There is usually a bit more to the identity conditions of these and other activities than simply aiming at a certain end, and this brings me to a second complication. For the purpose of simplicity, I have spoken about characteristic ends in abstraction from the *way* these ends are brought about. But our understanding of what it is, say, to cook also includes a somewhat determinate conception of the sorts of ways in which the ends of nourishment, etc. are produced.

This point can be illustrated by means of the possibility of deviant causal chains. *A* might demand that *B* cook for him and threaten *B* not with a beating but with depriving *B*'s children of their dinner. If *A* submits to this threat, then *A* cooks for the sake of nourishment—but clearly not in the characteristic way. Characteristically, cooking aims at nourishing the very person who is to consume its product, and to nourish that person *through* this act of consumption.

Here is another example.[4] I have suggested that, unlike cooking, self-defense is characteristically a way of avoiding a beating. Suppose, then, that *B* is a famous martial artist; *A* has overwhelmed and kidnapped *B*. *A* is armed and prepared to give *B* a beating, but all he wants is to experience a demonstration of *B*'s fighting skills. So he commands that *A* punch him (or whatever) and that *B* will beat him otherwise. If *A*

---

[3] Since neither is for money, defenders of the commodification of sex are right that there is nothing special about sex in this regard. But they are wrong insofar as they assume that there is no problem with the commodification of labor in general. See Sections 7–9 below for more on this.

[4] I owe the essentials of this example to Paskalina Bourbon.

submits to this command and punches *B*, he will be exercising his self-defense skill for the sake of avoiding a beating. But he will not be doing so in the characteristic way, for the characteristic way in which this type of activity achieves its end is obviously not by appeasing one's enemy.

There are bound to be more subtle examples still. It seems impossible, at least in most cases, to articulate the characteristic way in which an activity is a means to its end completely generally, so as to exclude all possible deviant causal chains. Moreover, as with the characteristic ends themselves, the way in which these ends are achieved admits of plurality and contestation—much more so, in fact. But again it would be a mistake to conclude that anything goes. At least usually, when we identify a person as doing one thing rather than another, our identification involves an understanding not just of what this activity characteristically aims at but also of *how* it characteristically aims at these ends, even if this understanding outstrips our ability to articulate it independently in terms of necessary and sufficient conditions.

2.3

But if we generally distinguish *kinds* of activity from one another in terms of their *characteristic* ends, why can we not cut out the middle-man and distinguish *particular* activities themselves in terms of their *actual* ends alone? Why does cooking for the sake of avoiding a beating not just become a novel form of self-defense?[5]

This is an important question for two reasons. First, it fundamentally challenges the distinction I have drawn. If the proposed redescription is accurate, then cooking

---

[5] Thanks to Jacob Butcher for pressing me on this point.

under coercion is not radically extrinsically motivated after all, and since this pattern of redescription clearly generalizes, there will then turn out to be no such thing as radically extrinsically motivated action at all. Second, more positively, understanding why this redescription does *not* in fact work will turn out to reveal and explain an important feature of radically extrinsic motivation, namely its paradigmatically interpersonal character.

It is possible to imagine cases similar to coerced cooking where one's actions really do amount to nothing more than an ingenious form of self-protection. For example, suppose a traveler in nature finds himself in an area infested with dangerous insects. He discovers that cooking a particular meal on his campfire, for reasons he does not understand, has the effect of keeping the insects at bay. Henceforth, he decides to "cook" this meal every night before he goes to sleep, so as to keep the insects at bay. He does not do this for the sake of nourishment; he does not like this particular "meal," has enough to eat already, and so does not eat what he "cooks." He cooks only for the sake of keeping the insects at bay. Here, strictly speaking, there is really no longer any reason to say that he is cooking at all. He is doing something which looks like cooking but is in fact a form of self-protection, or chemical warfare.

What is crucial about this case, though, is that it is a pure accident that "cooking" a certain "meal" happens to produce the relevant effect. Once our traveler discovers, say, that the same effect can be produced by simply burning one of the ingredients directly on the campfire, he can do that instead. Cooking *qua* activity which produces means of nourishment, etc. plays no essential role in what he is up to.

Contrast this with the case of coercion. If *A* credibly threatens *B* with a beating unless *B* cooks for him, and *B* does comply and cook for him, then the activity of

cooking, *under that description*—that is, *qua* activity for the purpose of nourishing someone—plays an essential role in what *B* is up to. For to the extent that *A* is an effective coercer and *B* is genuinely under his power, there is no activity other than, precisely, *cooking* which could serve as an alternative means for *B* to avoid a beating. In particular, replacing this activity with any alternative which resembles cooking as closely as you like except that it does not produce a means of nourishment for *A* would fail to bring about the end of avoiding a beating. This is not the case in the chemical-warfare example, where the desired effect can be, and indeed is, produced by another action which merely resembles cooking.

What is crucial about the case of radically extrinsic motivation, then, is that the radically extrinsic incentive tracks not just some accidental feature of the incentivized action but the action itself. It cannot do this, however, unless it is mediated by a representation of the concept of that action. For this reason, radically extrinsic incentives are essentially imposed by persons. When impersonal circumstances of nature make it the case that the movements which normally belong to a certain kind of activity happen, unusually, to produce a different desired effect, and when the movements are performed for the sake of this unusual effect, these effects are not radically extrinsic to the movements. For in these circumstances, the movements do not constitute the activities they normally constitute. It takes a *person* to ensure that not only the movements but the activities themselves become essential means to non-characteristic ends.[6]

---

[6] Perhaps it is also possible to motivate *oneself* in a radically extrinsic way; this would be a way of treating oneself as other.

3. RADICALLY EXTRINSIC MOTIVATION AS MERELY INSTRUMENTAL

In Chapter III, I argued that if an instrumental action has any intrinsic value, this intrinsic value must depend on the instrumental value of the action. For this reason, radically extrinsically motivated action is necessarily *merely* instrumental.

3.1

It does seem somehow obvious that to do something for the sake of a radically extrinsic end is to do it merely instrumentally and not for its own sake, but it is worth seeing why this must be true.

An instrumental activity like cooking need not in general be done *merely* instrumentally but can be done for its own sake, that is, on account of its intrinsic value. One might value cooking, for instance, not just on account of its production of a further good—means of nourishment—but also as the exercise of a valuable skill or as the making of a valuable social contribution. As I argued in Chapter III, Section 1.2, however, even insofar as one performs an instrumental action for its own sake, one necessarily has in view the concrete further end which makes the action the action that it is. For it is only by being internally connected with this further end that the action's intrinsic value belongs to the action as such—for instance, as cooking—and so can count as intrinsic to it. Thus, I argued that if *the exercise of a valuable skill* or *the making of a valuable social contribution* is an aspect under which cooking is *intrinsically* valuable, this can only be because the value of the relevant skill or social contribution depends on the fact that cooking has a certain instrumental value, namely the production of a means of nourishment.

But an agent who cooks without having this characteristic end in view also cannot have in view any intrinsic value which depends on this end, and so cannot be cooking for its own sake. For instance, one who cooks not for the sake of nourishing someone but merely for the sake of avoiding a beating *eo ipso* does not cook for the sake of exercising the valuable skill of cooking or for the sake of making the valuable social contribution of satisfying someone's need for food. Being disconnected from the ends which make cooking the activity it is, her end must likewise be disconnected from any value which cooking may have in itself, by virtue of being the activity it is.

Perhaps a more intuitive way to get at the same point is that the good which the radically extrinsically motivated cook is after really has nothing to do with cooking at all. The reason cooking happens to be an effective means to her end (of avoiding a beating) has nothing to do with cooking considered in its "concreteness"—in light of any of the properties to which one might point in order to distinguish it from other types of useful activity, such as housebuilding, journalism, or teaching. In this sense, as far as the agent herself is concerned, the means to her end is radically fungible. Considered in terms of its fittingness to her end, it is on a par with just about any other activity. But the deep lack of interest in one's concrete activity which this radical fungibility manifests is incompatible with doing it for its own sake.

Phenomenologically, the fungibility of radically extrinsically motivated action corresponds to an experience of especially profound alienation from what one is doing. The agent is indifferent to her own activity; she feels that there is actually no good reason to do it, that nothing speaks in its favor. She feels this way because radically extrinsically motivated action is not chosen on its own merits. It is not chosen for its

intrinsic satisfactions, but more broadly it is not even chosen on account of being a certain sort of productive, *instrumentally* valuable activity. For all that the agent is concerned, the world might be no worse (other things equal) if both this activity *and* its product were to vanish from the face of the Earth.

3.2

There might seem to be a problem here because it is possible to do something for two different reasons at once. In particular, it seems possible to do something both for the sake of some radically extrinsic motive and for its own sake at the same time. For example, someone who is commanded to cook a meal and who happens to love cooking could conceivably cook the meal both to avoid a beating and for its own sake. But if this is possible, then radically extrinsic motivation seems insufficient for *merely* instrumental action after all.

To be sure, we must take care to imagine just the right kind of case here, and this is harder than we might be inclined to suppose. It is not enough, if the agent is to count as performing the activity also for its own sake, that the activity in question happen to be intrinsically good, or even that the agent take it to be so. Its intrinsic goodness must actually be motivating her; it must be part of the explanation for why she does what she does. In particular, the agent must not merely be deluding herself into thinking that she is acting from an intrinsic desire, as a way of making a virtue of necessity. The intrinsic desire must be genuine.

In the real world, these conditions are likely satisfied only rarely. They are likely *known* to be satisfied more rarely still. Some Kantians say that one can never know the actual moral worth of one's actions in light of the inevitable presence of possible non-

moral motives. Similarly, one's power of practical self-knowledge is liable to be gravely impeded in conditions of bondage, with their attendant temptations of bad faith. None of this shows, however, that the relevant kind of practical overdetermination is impossible in principle. In principle, it is possible.

I think the correct thing to say about overdetermined action of this kind, though, is that it is done for its own sake under one aspect and merely instrumentally under another. This separation of aspects is justified because each aspect is strictly independent of the other; they constitute separate streams of practical reasoning which coincide only fortuitously.

Consider a person who cooks for its own sake, on account of the perceived intrinsic value of cooking, but not under coercion. As I have emphasized, the perceived intrinsic value of cooking depends on the perceived instrumental value of that activity. Insofar as she cooks for its own sake, that is partly because she also cooks instrumentally, for the sake of providing someone with something delicious and nourishing. While there are in a way two motives here, then, they are not independent; the one is parasitic on the other. The fact that she is motivated in both ways at once is not a mere accident.

Now suppose that in addition to this motive (or pair of motives), the same person cooks, on the same occasion, also in order to avoid a beating. As a radically extrinsic motive, this end has nothing to do with the other(s). The fact that the cook is motivated *both* in the first way(s) *and* in this radically extrinsic way is an accident. The radically extrinsic end thus constitutes a strictly separate, albeit simultaneous form of motivation. But on its own, as I have argued, it is a species of merely instrumental motivation. So there is an independent aspect of the cook's action under which it is

done merely instrumentally.

The opposite of merely instrumental activity, then, is not simply: instrumental activity which is done *in addition* for its own sake. For an instrumental activity to be more than merely instrumental in the relevant sense, it cannot be a mere accident that what is necessary for the sake of her further ends is precisely that which she wants to do for its own sake. But insofar as she is coerced to do what (as it happens) she loves to do, this can only be an accident. Insofar as she is coerced, then, she does act merely instrumentally.

4. THE UNFREEDOM OF COERCED ACTION

In this section, I put together many of the pieces I have assembled over the preceding sections and chapters. Given the account of freedom I defended in Chapter III, I argue that the radically extrinsic character of coercive incentives explains the unfreedom of coerced action. The proposed explanation satisfies many of the desiderata defended in Chapter I and solves several of the problems with extant theories discussed in Chapter II.

4.1

I have argued that coerced action as such aims at a radically extrinsic end. I have argued, further, that action motivated in this way is not just any old species of instrumental action but is necessarily a species of *merely* instrumental action: action done instrumentally and *not* for its own sake. But in Chapter III, I argued that a free action must be done for its own sake and not merely instrumentally. It follows that coerced action is unfree, and that its unfreedom is a species of the unfreedom of merely

instrumental action.

This unfreedom comes in degrees. In Section 4 of Chapter III, I argued that a merely instrumental action is unfree in proportion to the importance of its further end. In particular, then, we can say that a coerced action is unfree in proportion to the importance of its radically extrinsic further end. This principle will produce a clear ranking where there is a clear difference in importance of the goods in the balance. Thus, other things equal, the threat of killing someone is more coercive than the threat of giving them a beating, and the threat of stealing a poor person's possession is more coercive than the threat of stealing the same possession from a rich person. Any account of coercion should be able to explain these differences, so it is a theoretical virtue of my proposed account that it makes a ready explanation available.

More importantly, my account finally makes sense of the unfreedom of coerced action without at the same time effacing the agency of the coerced. It does this by making available a category of action which is neither a mere non-rational movement nor an *internally* impaired exercise of agency nor a full and unqualified exercise of agency. As the pursuit of the agent's own ends, a coerced action evinces genuine choice, unimpaired in respect of the agent's formation and pursuit of her desires or her determination and pursuit of her ends. As a species of merely instrumental action, however, it is unfree and represents only a qualified, privative exercise of agency. We thus resolve the paradox of unfree agency discussed in Chapter II.

Unlike the moralistic attempts to explain the wrongfulness of coercion, moreover, the account I am defending does not depend on the independent wrongfulness of the threat. Whether or not the mugger's threat to kill his victim, say, is wrong on

independent grounds, what makes it distinctively wrong as an act of *coercion*—that is, as a way of subjecting another's *will*—is that he issues the threat as a radically extrinsic incentive to action. And this feature of the threat is not grounded in its prior wrongfulness. We thus retain the correct order of explanation between unfreedom and wrongfulness.

By the same token, my account, unlike moralistic approaches to coercion, gives us a principled way of making sense of cases of coercion which seem not to involve an independently wrongful threat, such as the examples of blackmail and *quid pro quo* sexual harassment discussed in Chapter II. We can say that the threats made to the sexually harassed employee and the blackmailed political candidate distinctively wrong them not because the "offers" in question (not to fire the employee, not to publish the story) would be wrong anyway, but because when made in this way, they are coercive. They are coercive because they intentionally confront the agent in question with an incentive which is radically extrinsic to the incentivized action. The employer's incentive of employment bears no intrinsic relation to the sex he demands; the blackmailer's incentive of secrecy bears no intrinsic relation to the payment or service he demands.

This way of making sense of such cases turns out to bear a remarkable similarity to some of the explanations proposed by moralists. Most strikingly, as I mentioned in Chapter II, Wertheimer suggests that the problem with the blackmailer's threat is that it involves asserting one's rights in order to gain advantages which bear "no intrinsic

connection" with those rights.[7] The problem, in other words, lies in the lack of an intrinsic connection between the incentive (the blackmailer's threat) and the incentivized action (the object of the blackmailer's demand). This is more or less a version of what I have argued. The difference is that whereas the moralists' appeal to this sort of idea is *ad hoc*, meant only to deal with a special class of difficult cases, I have argued that it is fundamental to an understanding of coercion in general.

My proposed explanation of the unfreedom of coerced action is also an improvement on the psychologistic sort of explanation defended by theorists of autonomy such as Frankfurt and G. Dworkin. To be sure, the psychologistic approach, like its moralistic counterpart, contains a substantial core of truth. It is true, as Dworkin puts it, that we "resent acting for certain reasons," and in particular that we resent being motivated by coercive threats.[8] It is also true that this form of motivation is associated with an experience of alienation—a lack of "identification" with one's action, a sense that it is not one's own, that it does not wholeheartedly proceed from one's own will.

As I argued in Chapter II, what the psychologistic approach fails to recognize is that these moral and psychological truths do not explain the unfreedom of coerced action but are dependent on it. My account restores the proper order of priority, by explaining the unfreedom of coerced action not on the basis of its affective significance but directly on the basis of the motivational structure which characterizes coercion. It is

---

[7] Wertheimer, *Coercion*, 220. As we saw, other moralists also propose that there must be some restriction on the *kinds of reasons* for which employers should be able to fire their employees. Scanlon, *Moral Dimensions*, 83–87; Pallikkathayil, "The Possibility of Choice," 18–19.
[8] Dworkin, "Acting Freely," 377.

because this structure necessarily amounts to a form of merely instrumental motivation that coerced action is unfree, and it is because coerced action is unfree that coercion is a normal object of resentment.

4.2

So far, I have focused on vindicating the unity of freedom (as activity for its own sake), by showing that interpersonal unfreedom can be understood as a species of a more general type of failure (merely instrumental activity). To fully explain the distinctive wrongfulness of coercion, however, one also has to understand what distinguishes interpersonal from non-interpersonal forms of unfreedom. As I suggested in Chapter I, the problem to be addressed is that interpersonal unfreedom seems not merely to have a distinctive etiology or to be of distinctive moral significance but also to be distinctive, as one might put it, *qua unfreedom*. That is, it seems to involve a special—and especially profound—*way* of having one's will impeded or compelled. The challenge is to make sense of this thought. My attempt to do so has two parts; I describe the first here and the other in the next section.

I have argued that free activity is activity for its own sake. This means that merely instrumental action is unfree, whether or not any human agency is the cause of its necessity. Consider, once again, two ways of losing a kidney. One might undergo transplant surgery in order to donate one's kidney to a relative dying of kidney disease. Or one might undergo the same surgery because one's relative has been kidnapped and the captors are demanding one's kidney as ransom. In both cases, we can suppose that one does the action not on account of its intrinsic value but solely for the sake of a further end. One acts unfreely in both cases. But only the second case involves radically

extrinsic motivation, since the characteristic point of kidney transplant surgery is the life and health of the organ's recipient, threatened with death by kidney disease, and not the life of the kidnapped relative, threatened with violent death by shooting.

What difference does this make to one's unfreedom? To begin with, as I argued in Section 2, the radically extrinsic incentive in the second type of case is characteristically imposed by another person, since the usefulness of the means—undergoing surgery—for the end—saving someone from being shot—depends on someone's *representation* of the means–end nexus. By contrast, the means–end connection in the first type of case—the usefulness of undergoing surgery for the end of saving someone from kidney disease—is there anyway, independently of anyone's representation of the nexus. Even if the agent had been *persuaded* by someone to a undergo the surgery for the sake of saving the relative's life, this form of interpersonal influence would not have brought the necessity of undergoing surgery (for the end of saving a life) into existence but would merely have pointed the agent to a necessity which is there anyway.

The distinctiveness of radically extrinsic motivation goes beyond its distinctive etiology, however. In the first type of case, the reason the action is unfree has to do with its concrete content. For the sake of argument, I am assuming that kidney surgery is not the sort of activity that has intrinsic value. It is unpleasant and risky, and for the crucial part of the process one is unconscious; if we could manufacture organs synthetically instead of cutting them out of people, we would simply do that instead, without thereby sacrificing anything of value. If undergoing kidney surgery in the first type of case is unfree, in other words, it is unfree by virtue of considerations concerning the concrete activity—specifically, whether this activity has any intrinsic value. (For the

purpose of this discussion, I am ignoring the difference between actual and merely apparent intrinsic value.)

In the second, characteristically interpersonal case, on the other hand, the unfreedom has nothing to do with the content of the activity in question. If undergoing kidney surgery *were* the sort of thing that was intrinsically valuable, or if the kidnappers had instead demanded that one do something else which happened to be intrinsically valuable, one's activity *still* would not be free. The unfreedom in the interpersonal case is distinctively a function not of the content of the action but of the form of motivation in play. That is to say, coercion turns *whatever* it is that the agent is coerced to do into a mere means, irrespective of its content and irrespective of any intrinsic value which it may possess. It is impossible to act for its own sake insofar as one is motivated in this way. This is how interpersonal unfreedom differs, as a *way* of being unfree, from non-interpersonal unfreedom. We could label the difference by saying that interpersonal unfreedom is *formal* rather than *material*.

We can infer two corollaries from the difference. First, whereas some substantive ethical reflection about what is and what is not intrinsically choiceworthy is needed in order to diagnose unfreedom of the non-interpersonal kind, no substantive ethical reflection of this kind is needed to diagnose unfreedom of the interpersonal kind. Second, whereas the non-interpersonal unfreedom of an activity can in principle be eliminated by transforming the activity's content so that it becomes intrinsically valuable (I will discuss this more in Chapter V), the interpersonal unfreedom of an activity cannot be eliminated by transforming its content but requires changing the form of its motivation, which of course requires transforming the interpersonal relations

in which it is embedded.

5. THE SUBVERSION OF PRACTICAL REASON

A second feature of radically extrinsic motivation which sets it apart from merely instrumental motivation in general is the way in which it constitutes a distinctively impeded exercise of the power of practical reason. Specifically, I argue in this section that to do something for the sake of a radically extrinsic end is not to do it because one takes oneself to have good reason to do it. In this way, I show that we can now repair the Kantian view of freedom and coercion criticized in Chapter II. While the amended view still has limitations when it comes to the understanding of non-interpersonal unfreedom, it can shed considerable light on what is special about interpersonal unfreedom.

5.1

When one acts for the sake of an end that is a characteristic further end of one's action, one chooses one's action on its merits. For instance, when one cooks a meal in order to nourish someone, one's end speaks in favor of that very action. In this sense, one takes oneself to have good reason to do just what one chooses to do. This is true even if someone else has persuaded one, in the ordinary rational way, to cook for them, perhaps by making it clear that they need the food but are for some reason unable to cook the meal themselves. Even then, where one acts in some sense under the influence of another's will, one's decision is ultimately made on the basis of one's own assessment of the merits of cooking: one cooks a meal because one takes there to be good reason for cooking a meal.

The same is true of actions which are unfree but whose unfreedom does not have its source in another person. In the case of a kidney donation to a dying relative, for instance, we can assume that the action is not done strictly for its own sake but merely as a means to a further end. Nevertheless, it *is* done on its own merits, on the basis of the further end that belongs to it as the kind of action it is. The health and life of the organ recipient is exactly what transplant surgery is good for. There is thus good reason to choose just this action, and that is why one does it.

By contrast, in cases of interpersonal unfreedom, where one does something just for the sake of a radically extrinsic end—such as cooking a meal in order to avoid a beating—there is a sense in which one does not take oneself to have good reason to do what one is doing. Again, someone's need for nourishment is a good reason to cook them a meal. The threat of a beating is a good reason to practice self-defense. But the threat of a beating is not a good reason to cook a meal; it no more speaks in favor of cooking a meal than it speaks in favor of performing any other action.

In fact, in one's own assessment of cooking on its own merits, this action may well have nothing at all to recommend it. It would provide nourishment for someone, but one may have no interest in nourishing this obviously unpleasant individual. In this way, we can see that the radical fungibility of one's action which I discussed in Section 3, and the especially profound sort of alienation associated with this fungibility, is a distinctive feature of the interpersonal kind of unfreedom. Only where one's further end is radically extrinsic to one's action does one have no interest at all in the action itself, even considered as an inherently useful thing to do, that is, as the sort of action which characteristically aims at the good one is pursuing.

There is a distinctive sense, then, in which the coerced agent does not take herself to have good reason to do what she is doing. Indeed, she may even be acting *in spite of* her conviction that there is no good reason to do it, and that there is good reason *not* to do it. And this explains the sense in which an agent's power of practical reason is distinctively subverted in interpersonal unfreedom. If practical reason is (among other things) a power to act on the basis of one's own determination of what there is good reason to do, then one exercises this power only in a privative, qualified way insofar as one acts for the sake of a radically extrinsic incentive. One's action may be based on one's judgment that it is the thing to do in the circumstances, but it is not based on one's judgment that there is good reason to do it.

It is worth noting that this feature of radically extrinsic motivation also explains why *resistance* to coercion, unlike submission to coercion, can be free, at least as far as the interpersonal dimension of freedom is concerned. Using a martial art in self-defense, for instance, is the sort of action which *does* aim at avoiding a beating. So in defending oneself against a would-be coercer's assault, one does just what one takes oneself to have good reason to do, because one takes oneself to have good reason to do it.

5.2

The explanation I have just given of what is special about interpersonal unfreedom puts us in a position to repair the Kantian conception of freedom and unfreedom which I criticized in Chapter II. The basic Kantian idea was that coercion wrongs another because it is a way of using her that does not respect her status as a rational agent, that is, a being with the power to make up her own mind about what she has good reason to do. Korsgaard, to reiterate, writes:

> To treat others as ends in themselves is always to address and deal with them as rational beings. Every rational being gets to reason out, for herself, what she is to think, choose, or do. So if you need someone's contribution to your end, you must put the facts before her and ask for her contribution.[9]

Deception and coercion undermine another's autonomy and wrong her, on this view, because they are ways of failing to treat the other as a rational being in this way.

The trouble was that it turned out to be difficult to make sense of what it could be to prevent another from properly exercising her ability to act on the basis of her own judgments of her reasons for action, short of completely effacing her agency. If someone's cooperation is obtained by means of a coercive threat, it would seem that it still remains up to her to determine what she has reason to do; it is merely that these reasons have been changed.

My account of what is wrong with coercive incentives puts us in a position to vindicate at least the spirit of the Kantian picture. What we need is to make sense of the possibility of a *genuine* but *impeded* exercise of the power to act for reasons. Action for the sake of a radically extrinsic end fits this description. As I have argued, to act for the sake of a radically extrinsic end is to exercise one's practical reason, but only in a way which falls short of acting on the basis of a judgment to the effect that one has good reason to do just what one is doing. To get someone to do something by means of a radically extrinsic incentive, then, is to use her in a way which fails to respect her as a rational agent, understood (in part) as someone with the power to act on the basis of her own judgment of what she has good reason to do. To influence her will in a way

---

[9] Korsgaard, "The Right to Lie," 142.

consistent with respect for this status, one would indeed have, as Korsgaard says, "to put the facts before her and ask for her contribution."

This way of fleshing out the Kantian view puts us in a position to reply to some of the objections leveled against it. Most fundamentally, it allows us to meet challenges of the sort I myself pressed in Chapter II. Buss, for instance, points out that a victim of manipulation "must act for her own reasons if she is to act at all; and so she must act for her own reasons if she is to do what [her manipulator] wants her to do."[10] She concludes that the Kantian view is mistaken: "if manipulating and deceiving someone are generally wrong, this cannot be because these modes of interacting with other rational agents […] deprive these people of their autonomy."[11]

But we can now see that this is too quick. A manipulator could undermine his victim's autonomy not by completely suspending her ability to act for her own reasons, but merely by inducing her to act on something less than a *good* reason. And it is the concept of a radically extrinsic end—really a defective kind of end—which gives content to this notion. Admittedly, my defense requires dropping some of the most extreme Kantian formulations, which imply that the coerced agent does not really act for reasons at all. These formulations are Buss's typical target, and I agree that they are implausible.

My reinterpretation also clarifies the special significance of the fact that coercion involves incentives meant to influence another's actions. This helps fortify the Korsgaardian thought against another type of objection, sometimes raised by other

---

[10] Buss, "Valuing Autonomy and Respecting Persons," 213.
[11] Ibid., 217.

Kantians, which seems to me a smidgen unfair. Pallikkathayil, for instance, argues that the ideal of fully consensual rational cooperation is unreasonably demanding. She points out that there are ways of "using" another's "agency" without her consent which do not wrong her. For instance:

> Suppose that you are trying to decide whether it is cold enough outside to wear a coat. You look outside your window to see whether people on the street are wearing coats. Here you use their actions (walking around with or without a coat) to further your end (determining what to wear).[12]

Since this sort of conduct is not wrong, "using" others' "agency" without their consent cannot be wrong as such.

As Pallikkathayil also acknowledges, however, this way of "using" others is not a way of "intervening in their activities to get them to do what you want. Indeed you are simply making use of what they are already doing."[13] But in that case, this way of "using" others does not depend on these others acting for bad reasons, or exercising their agency on account of anything other than their own judgment of what they have good reason to do. So at least on my interpretation of the ideal of rational cooperation, we need not say that there is anything wrong with this form of "using" others.

5.3

I think some of the ideas I have been developing here also find expression in the work of Barbara Herman and of A.J. Julius. It seems to me, however, that my own approach more directly captures what is wrong with coercion from the coercee's point of view.

Explaining what is wrong with deception, Herman writes:

---

[12] Pallikkathayil, "Deriving Morality from Politics," 127.
[13] Ibid., 129.

> The maxim of deception places the grounds of choiceworthiness of the victim's reasons in the deceiver's will. The victim cannot take the deceiver's action as good for the reasons that make it good (in the eyes of the deceiver) all the way down.[14]

In the background of this explanation is an ideal of interpersonal interaction: namely that one should only attempt to induce another to do something by getting her to do it for the reasons that make it good. It would seem that this ideal can also help us make sense of what is wrong with coercion, in a way which allows that the coercee might be acting for reasons and which thus does not simply efface her agency. The point is that, *from the coercer's point of view*, the reasons on which the coercee acts (to avoid a beating, for instance) are not the reasons on account of which the action is choiceworthy. So as far as the coercer is concerned, then, he is inducing someone to perform an action on the basis of something other than "the reasons that make it good."

This seems right to me, as far as it goes. However, it does not quite manage to capture what is wrong with coercion *from the coercee's point of view*. After all, might not the coercee have *her own* reasons for doing what the coercer wants her to do—for instance, to avoid a beating? For all that has been shown, these might be good reasons from her point of view even if they are not what makes the action choiceworthy as far as the coercer is concerned. Tellingly, Herman's conclusion about the parallel case of deception is:

> The moral problem is not *in* the will of the deceived, but in the maxim of the deceiver. Fault is therefore present independent of injury, regardless of whether the deceived person minds acting for the deceiver's end.[15]

---

[14] Herman, "Leaving Deontology Behind," 229.
[15] Ibid., 230, emphasis hers.

But at least for the case of coercion (as opposed to deception), this would be an uncomfortable position. As I argued in Chapter II, it is difficult to appreciate the wrongfulness of (attempted) coercion if such treatment (when the attempt is successful) does not *actually* adversely affect the other's will from her own point of view.

Julius's account of what is wrong with coercion rests on what he calls the "independence principle," which enjoins us not to try to induce another to do something "except by helping the person who performs the action to do it for a reason that favors the action independently of what is done to produce it," that is, independently of the inducement.[16] The ideal of interpersonal interaction in the background here, too, is premised on respecting another's status as a practical reasoner, understood now as a being with the power to act for the reasons she has *anyway*, independently of anyone's attempts to influence her will. Attempted coercion violates this ideal because a coercive incentive is patently not a reason for action which one has anyway.

I do not disagree. In fact, in arguing from the independence principle, Julius focuses (at some level of description) on what amounts to the same feature of coercive incentives which I have singled out as crucial, namely that by virtue of which it makes sense to call them "incentives" (see Section 1.3 above). An incentive motivates an action by artificially attaching a good to it which does not attach to the action *anyway*, a good which it would not otherwise produce. An advantage of Julius's approach vis-à-vis my own, moreover, is that it is more formal than mine, apparently dispensing with the

---

[16] A. J. Julius, "The Possibility of Exchange," *Politics, Philosophy & Economics* 12, no. 4 (2013): 363.

needlessly contentious process of determining whether a given end is "characteristic of" or "radically extrinsic to" a given action.

The question, though, is whether, in shedding this ballast, we are not actually forgoing something quite helpful. The flipside of the formality of Julius's approach, as I see it, is that the independence principle, like Herman's account, does not directly illuminate what is wrong with coercion from the point of view of the coercee. In particular, it does not tell us why, when the principle is violated and a coercer does induce a coercee to act on the basis of some non-independent consideration, this consideration should not—now that it has come into being—count as a perfectly good reason for the coercee.

I think that it is not in fact a good reason. But this claim stands in need of explanation and justification. Otherwise, we still lack an explanation of how coercion actually impedes its victim's power of practical reason, and by the same token we risk losing our grip on why we should believe the independence principle in the first place. One wants to say: We should not induce an agent to act only on the basis of non-independent reasons at least partly *because* such reasons are bad reasons, from the agent's own point of view. But we need an explanation of what this means and why it is true, and I do not see how the independence principle can provide it.

My account of radically extrinsic ends and what is wrong with them answers the questions raised by both Herman's and Julius's accounts. To do something for the sake of a radically extrinsic end is essentially not to do it on its own merits, or on the basis of reasons which speak in favor of that very action. A coercive incentive, because it fits this structure, therefore fails to give the coercee a good reason—the reason, or even *a* reason,

which makes the action choiceworthy—from the coercee's own point of view. And that is the problem with setting out to influence another's will in this way.

6. THE WRONGFULNESS OF COERCION

Having explained the unfreedom of coerced action, we are now also in a position to explain the *pro tanto* wrongfulness of coercion.

6.1

To coerce another is intentionally to induce her to do (or refrain from doing) something by attaching a radically extrinsic incentive to the relevant action (or omission). If the attempted inducement is successful, it will necessarily involve the induced party acting merely instrumentally—that is, unfreely. This connection between the coercer's act and the coerced agent's unfreedom is not accidental. The coercer does not merely happen to bring it about that the other exercises her agency in a merely instrumental way. Rather, as I argued in Section 4, the very form of motivation which the inducement involves is merely instrumental. To coerce another is thus to use her agency in a way that is *essentially* incompatible with its free exercise. Coercion is essentially the mere instrumentalization of another's agency.

The same conclusion emerges from the discussion of Section 5. As I argued there, coercion distinctively impedes another's exercise of practical reason by inducing her to act for something other than her own judgment of what she has good reason to do. This offends against her status as a rational agent. It is a way of making use of the other's practical reason which, if successful, is incompatible with her free exercise of that

power.

Once again, the contrast with rational persuasion is instructive. One may get another to help one attain one's own ends by persuading her that she has good reasons to do so. For instance, one may convince her that one's project advances certain further ends which she greatly cares about. If she agrees to help for this reason, there is *some* sense in which one "instrumentalizes" her agency, or "uses" her as a means to one's ends. But since this way of "using" someone does not as such involve the imposition of a radically external incentive, it is not as such incompatible with her free exercise of agency. Her action may still be unfree, if it happens not to be choiceworthy for its own sake, but the ground of this unfreedom will not be the interpersonal relation between persuader and persuaded.

In a distinctively strong sense, then, to coerce another is to use her as a mere means rather than as a person capable of self-determination, i.e., activity for its own sake. But to use another as a mere means rather than as a person in this sense is thereby to wrong her. It is to violate her claim against others not to have her agency used in ways essentially incompatible with its free exercise. We can thus give substance as well to the Kantian formula that coercion is a way of using another as a mere means—but without having to pretend incoherently that coercion somehow suspends or destroys the coerced agent's agency or personhood altogether.

The normative foundation of my proposed explanation is nothing but each person's generic right to freedom, which gives rise to a claim against each other person not to coerce her. This, too, seems in keeping with the spirit of Kant, who, as we saw in Chapter II, defines the very concept of right in terms of freedom and regards the right to

freedom as foundational for all other rights. The moralistic appropriations of Kant were unable to treat this right as truly foundational, since they had to appeal to other, independent rights in order to ground or even give any content to the right against coercion. My approach reinstates what seems to be a more promising order of moral and conceptual priority from the point of view of the theory of rights as a whole.

6.2

Evidently, the ideal of interpersonal interaction which animates this account of what makes coercion wrong is very demanding. A range of putative counterexamples suggests itself. Perhaps the most theoretically and politically important class of cases are cases of offers (as opposed to threats); I will deal with these in Sections 7–9. Here, I will first address two other classes of putative counter-examples.

The first class comprises coercive threats which have the putatively wrong-making features I identify but which do not seem wrongful. One kind of example is (some) coercion by parents of their children: for instance, a parent's threat to take away the child's phone unless the child goes to bed. Another example is (some) coercion by the state of its citizens: for instance, the threat to prosecute and fine or incarcerate citizens unless they obey the law.

Intuitively, there is every reason to consider these threats coercive: they are ways in which one agent gets another to do something against their own will; the apparent coercer is in a position of nearly absolute power over the coercee in certain regards; and the coercee does not even have a meaningful possibility of exit from the coercive relationship. More to the point, the operative incentive in both cases appears to be radically extrinsic to the incentivized action: keeping one's phone has nothing to do

with going to sleep, and avoiding incarceration has nothing to do with paying one's taxes. Yet one might reasonably think that not every such case of coercion is wrongful. So it seems that giving another a radically extrinsic incentive is not always wrong. Is this a problem for my account?

Not really. By general consensus, the relation of parents (and other caregivers) to the children in their charge, and of the state to its citizens, is exceptional and licenses forms of treatment which would normally be indefensible. The fact that the general wrongfulness of coercion is sometimes defeated in these special contexts, however, does not show that it is a morally irrelevant consideration even in these cases. If, for instance, there is a reasonable alternative, non-coercive way to persuade a child to go to bed, then the general wrongfulness of coercion would seem to constitute an overriding reason to choose that alternative instead. Moreover, it may turn out in any case that children are more capable of understanding, judging of, and acting on reasons than they are commonly given credit for. Given the way in which radically extrinsic incentives disrespect persons' status as agents, this possibility gives us another reason to be very careful in condoning the use of such incentives on children (quite apart from the evidence that they are likely to be counterproductive).[17]

Some of the same kind of reasoning applies to the case of state coercion. The details will depend on something approaching a complete political theory, which is obviously beyond the scope of this dissertation. All I will say here is that the

---

[17] Deci *et al.*'s comprehensive meta-analytic study has shown that the general tendency of extrinsic rewards to undermine intrinsic motivation is especially pronounced in children. Deci, Koestner, and Ryan, "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivations," 641, 657.

justification for coercion by the state would seem to be strongest when such coercion is in the service of supporting each person's right to freedom itself (a type of justification which may sometimes also be available in the case of the coercion of children).

6.3

The second class of putative counter-examples comprises threats which have the putatively wrong-making features but which seem neither coercive nor wrongful. Specific examples will vary greatly, but here is one from the literature which I take to be representative of the sorts of worries which my view might provoke. Stephen White writes:

> If a coach threatens to make her players run extra laps if they are late to practice, this is presumably because she is not confident they are sufficiently motivated by the independent reasons favoring punctuality. But it hardly seems that, in introducing this new incentive to arrive by the start of practice, the coach thereby wrongs the members of her team.[18]

Unlike the cases considered above, one may be inclined to think that this sort of interaction does not even rise to the level of coercion—but in any case, it certainly seems like it could be morally innocent. To drive the point home, notice that the coach's threat seems altogether more troubling (and undeniably coercive) if what she threatens to do if the players come late is to slash the tires on the their cars.[19] In both cases, the coach's incentive (unlike "the independent reasons favoring punctuality") seems to be radically extrinsic to the incentivized action. Yet only the second case seems to involve a genuine moral wrong.

---

[18] White, "On the Moral Objection to Coercion," 212.
[19] This is an adapted version of another of White's examples. Ibid., 206.

One slightly odd feature of the first scenario is that the threat to "make" the players run extra laps is really a kind of meta-threat, since the coach can only "make" the players run laps, at the end of the day, by means of some further threat, such as the threat of barring them from joining practice that day, or of kicking them off the team altogether. Putting this complication aside, though, it seems to me that there are two ways of imagining the scenario, and neither version undermines my account of the wrongfulness of coercion.

First, let us imagine the best-case scenario, where the coach's actions are indisputably morally unproblematic. In this scenario, it seems to me that if the coach's pronouncement plays any role at all in motivating the players, it functions not as a radically extrinsic incentive but as a symbolic reinforcement of the meaning of coming or failing to come on time. To attach a penalty to tardiness is not, in the ideal case, to *incentivize* punctuality but rather to send the *message* that coming on time is an important part of being committed to the team and that in failing to do so one lets the team down.

To see this, imagine that the penalty is not being made to run extra laps, which might be exhausting and unpleasant in itself, but merely being made to do a few push-ups. Even assuming that fit athletes would not feel *in the least* perturbed by the prospect of doing a few push-ups in itself, they might well find this "threat" just as motivating. The reason, I submit, is that it has the same expressive, symbolic significance, which is what, in a "good" version of the story, is doing the real motivating. But to be motivated by the symbolic significance of coming on time is not to be motivated by anything radically extrinsic to that action; in the context of a team sport, that really is part of what

the action is for.

Given this interpretation, moreover, it makes sense that the "threat" of being made to run extra laps (or to do push-ups) would seem unproblematic in comparison with the threat of having one's tires slashed. Given the context, it is difficult to imagine how having one's tires slashed could have the same expressive significance that running extra laps can have. Whereas running extra laps is already a normal part of the team's training, having one's tires slashed is decidedly not. It is thus unclear how having one's tires slashed could serve the symbolic function of expressing one's commitment to the team and atoning for one's lapse.

That said, there may be a second conceivable version of the lap-running scenario. It is perhaps *possible* (though not that straightforward to imagine) that the threat of being made to run extra laps might play the role of a radically extrinsic incentive, simply as an unpleasant activity with no particular expressive function. But *if* this really does describe the motivation of the players, and if this is the way in which the coach intends to motivate her players, then it hardly seems obvious to me that what she does is morally unproblematic.

Various factors are liable to interfere with our "intuitions" here. For one thing, we may still be tempted by the first, "good-case" interpretation of the scenario. There is also the relative triviality of the stakes: presumably running a few extra laps would not endanger the players' health, for instance. Then there is the fact that, whatever the *moral* status of the coach's behavior, it is not, and presumably should not be, against the *law*. Finally, there is the fact that the coach presumably does what she is doing not merely in the service of her private ends—such as her own glory—but for the sake of an end

which the players share.

If we do not allow ourselves to be distracted by these factors, we may see the moral facts more clearly. The bottom line is that if the coach really does attempt to present the players with a radically extrinsic incentive, rather than presenting them with good reasons to come on time, something is surely amiss in this team, not just from a sporting point of view but morally. The coach fails to respect her players as independent agents. In a word, she treats them like children—but even if they happen to *be* children, the coach–athlete relationship does not normally give rise to the special justifications of coercion to which the parent–child context can give rise. The coach should not relate to her players in this way, and in choosing to do so she wrongs them, in the sense that they are entitled to demand that she stop.

7. THREE KINDS OF EXCHANGE

The feature of coercive threats which I have identified as the ground of the unfreedom of coerced action and of the wrongfulness of coercion is their radically extrinsic character. As I noted in Section 1, however, this is a feature which they share with other types of incentive, such as certain kinds of offer. A question therefore arises concerning the nature of the difference between threats and offers, and between coercion and a certain kind of exchange. In the rest of this chapter, I discuss what these forms of incentive and interaction have in common, and how they differ. I begin in this section by distinguishing three kinds of exchange so as to single one of them out as uniquely and essentially problematic.

7.1

Exchange is a form of interaction characterized by reciprocity. When we exchange with each other, there is some way in which you do your part and I do mine, and there is some deliberate relation of dependence between our respective contributions. There are several importantly different types of reciprocity which fit this description. They fall into at least three classes: the reciprocity which belongs to the exchange of favors, that which belongs to cooperation, and that which belongs to what I will call pure exchange. My arguments in this chapter bear directly only on pure exchange. To get this category into view, we need to distinguish it clearly from the other two.

One familiar form of what could loosely be called exchange is the exchange of gifts or favors. There are many different versions of such interactions, but broadly speaking they can be motivated in one or both of two ways. A gift can be given as a way of returning a previous favor. The reciprocity which belongs to a gift given in this way is that of gratitude. It is essential, of course, that the gift not be given for the sake of receiving anything in return; its reciprocity is backward-looking, not forward-looking.

A more forward-looking version of the exchange of gifts is also possible: one may exchange gifts as a way of building or maintaining a relationship of mutual trust or care. There is a kind of reciprocity here in the sense that the relevant (symmetric) kind of relationship can only be successfully maintained in the long run if both parties make contributions. Still, however, neither party's gift is motivated by the end of getting the other's gift in return. The end, which is shared by both parties, is rather the relationship itself (even if the relationship is in turn instrumental to further ends).

7.2

A second broad category of reciprocal interaction is cooperation. The paradigm case of cooperation is the pursuit of a shared end, such as the maintenance of a household or the victorious fighting of a war. Cooperation normally involves some division of labor in which each does their part in the service of the shared end, and each makes this contribution in the expectation that the others will make theirs.[20] The reciprocity manifested in the expectation that others will do their part is normally essential, as it is normally the case that any one individual's contribution on its own would not bring about the shared end. That said, it would probably be stretching the ordinary meaning of the term to call all cooperation of this paradigmatic kind "exchange" at all.

There is, however, a specific form of cooperation which certainly does qualify as exchange; we could call it "cooperative exchange." Julius suggests the following possibility:

> An exchange between two persons, I imagine, consists in each person's doing what will help the other because she believes that the pair of these actions is jointly recommended to the two of them as a pattern in which each helps the other. […] If each person does her part because she believes that the two are jointly recommended, each part happens because its author takes it to bear this significant relation to the other.[21]

In other words, in this kind of interaction, each party satisfies the other's need in the pursuit of the shared end of *satisfying each other's needs*. It is the distinctive character of this shared end which sets cooperative exchange apart from cooperation in general. The

---

[20] For an account of the sort of thing I have in mind, see Ben Laurence, "An Anscombean Approach to Collective Action," in *Essays on Anscombe's Intention*, ed. Anton Ford, Jennifer Hornsby, and Frederick Stoutland (Cambridge, MA: Harvard University Press, 2010).
[21] Julius, "The Possibility of Exchange," 369.

shared end is, as Dan Brudney puts it, "internally oriented": it consists in a way of relating to each other in a certain way, and not in some good above and beyond this social relation (such as winning a war).[22] It is, moreover, an "intertwined" shared end, in the sense that it cannot *in principle* be attained except "with and through others."[23] It would not be attained, for instance, if one or the other party's need were satisfied by mana rained down from the heavens.

By virtue of these features of the shared end, cooperative exchange can be said to involve an even stronger form of reciprocity than generic cooperation. In cooperation in general, the instrumentality of each individual's contribution to the shared end depends on others doing their part too, but in the case of cooperative exchange, others doing their part too is a *constitutive* part of the end and not merely causally necessary for it. For this even deeper reason, then, each party to a cooperative exchange does her part in the expectation, and on the condition, that the other will do hers.

7.3

All of the forms of exchange discussed so far involve a kind of reciprocity, a certain conditionality of one party's action on that of the other. It is an important fact, however, that in none of these cases can either party to the exchange be said to make her contribution *for the sake of* inducing the other's contribution in return. In the case of returning a favor out of gratitude, the action is backward-looking and so not done for

---

[22] Daniel Brudney, "Community and Completion," in *Reclaiming the History of Ethics: Essays for John Rawls*, ed. Andrews Reath, et al. (Cambridge: Cambridge University Press, 1997), 397.
[23] Ibid., 397–398.

the sake of receiving a new favor in return. In the case of forward-looking gift exchanges, each party's end is not receiving the other's gift in return but rather the shared end of building or maintaining a certain kind of relationship. The same is true of cooperative exchange.

But of course there is a kind of exchange, which we might call "pure exchange" or "mere exchange," where each party does make her contribution in order to induce the other's contribution in return. I cut your meat so that you will bake my bread; you bake my bread so that I will cut your meat. In (a pure case of) this kind of interaction, there is no shared end. In particular, we do not aim at the shared end of what I called "cooperative exchange," a relationship characterized by the mutual satisfaction of needs. If baked bread had simply materialized for me like mana, the end which I pursue in cutting your meat would equally have been attained.

Pure exchange is of course the form paradigmatically taken by market transactions. As Adam Smith famously observes:

> It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages.[24]

To be sure, the pursuit of "self-interest" must be constrained by certain norms of fidelity and non-aggression if people are to be able to trust each other sufficiently to enter into agreements in the first place.[25] Moreover, as is frequently pointed out, the good which

---

[24] Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (Chicago: University of Chicago Press, 1976), Book I, Chapter II, 18.
[25] Debra Satz, *Why Some Things Should Not Be for Sale: The Moral Limits of Markets* (Oxford: Oxford University Press, 2010), 28–29.

someone seeks from a market exchange may well be wanted for altruistic reasons and is anyway likely to be shaped by an ethical value system.[26] What Smith means, however, and what he is right about, is that the proximate aim of one's own contribution to a pure exchange is not the way in which this contribution will satisfy the other's need or desire. One is not motivated by the good which one's own contribution will directly do but by the good which one hopes to obtain from the other party's contribution (which may be as altruistic as you like).

At least this is true of a pure exchange as such. Real-life market exchanges can often be expected to involve a complicated mix of motives, which will blur the boundaries, in practice, between the various categories of exchange which I have distinguished in theory. For instance, in less anonymous settings, one aim of a transaction may be to build relations of trust with a particular vendor or customer, which will put the particular transaction somewhere in between pure exchange and the exchange of gifts. Similarly, particular market exchanges in the real world might approximate the model of cooperative exchange if the parties happen to be motivated in part by the way in which one's own contribution to the exchange will satisfy the other's need.

That said, in a large market society where transactions tend to be fairly

---

[26] See, e.g., Amartya Sen, *Collective Choice and Social Welfare*, Expanded ed. (Cambridge, MA: Harvard University Press, 2017), 50–51; Janet Radcliffe Richards, "Nephrarious Goings On: Kidney Sales and Moral Arguments," *The Journal of Medicine and Philosophy* 21 (1996): 392; Jason Brennan and Peter Martin Jaworski, "Markets Without Symbolic Limits," *Ethics* 125 (2015): 1058–1060; Barry Maguire and Brookes Brown, "Markets, Interpersonal Practices, and Signal Distortion," *Philosophers' Imprint* 19, no. 14 (2019): 5–6.

anonymous and where individual producers and consumers cannot satisfy their own needs except by transacting with others on the market, we are entitled to assume, as a reasonable approximation, that the parties are generally motivated by self-interest *in the relevant sense*, i.e., that their contribution to exchange is motivated not by the fact that it is a means to the other's ends but by the fact that it enables them to attain their own (perhaps altruistic) ends. Indeed, the economic theories which explain the well-documented productive and allocative efficiency of markets would seem generally to depend on this assumption. For instance, the tendency of competition to increase productivity depends on the assumption that consumers treat different producers as interchangeable, so that they are ready to switch allegiance to a new producer should the latter come to offer the same product at a lower price.

Payment for labor, in particular, similarly plays a variety of functions in the real world. In addition to functioning as an incentive, it can function, for instance, as a symbol of recognition or esteem. It is worth bearing in mind, though, that the coincidence of these functions is entirely contingent. In a different form of society, the same expressive or symbolic function could be played by something else. Indeed, if (as I will argue) there is something problematic about the incentive function of payment for labor, then this would seem to give us reason to transform the social world in such a way that it will no longer give rise to this particular form of expression of social recognition or esteem.[27]

---

[27] Brennan and Jaworski defend markets against "semiotic" objections (to the effect that market exchanges have an objectionable expressive or symbolic significance) by arguing that, given the consequentialist arguments in favor of markets, the semiotic objections simply give us reason to transform our practices of meaning-giving so as to divest

In any case, for the purpose of the following discussion, I will directly consider only clear, unmixed cases of the various types of exchange, on the premise that these represent useful idealizations of real-world interactions.

8. COERCION AND PURE EXCHANGE AS SPECIES OF EXPLOITATION

In this section, I argue that pure exchange is morally problematic in fundamentally the same way as coercion.

8.1

Of the kinds of exchange I have distinguished, only the last—pure exchange—involves radically extrinsic motivation.

Consider the exchange of gifts or favors. One returns a gift or favor "in kind": by giving the other a gift or doing the other a favor also. But one's action counts as the giving of a gift or the doing of a favor only because it satisfies the other's need or desire, and is done for this purpose. For example, if one knows that the other would enjoy a good meal, one might return their previous favor by cooking them dinner. But here the end and the means are internally connected. One chooses to cook a meal, rather than to perform some other action, because it satisfies the other's need for exactly the good

---

market exchanges of the relevant significance. Brennan and Jaworski, "Markets Without Symbolic Limits." I am pointing out, in effect, that this sort of argument cuts both ways. Given the deontic arguments *against* markets, we have reason to transform those practices which invest payment for services with valuable social meanings. (On the other hand, Brennan and Jaworski also fail to take seriously enough the possibility that some of the symbolic significance of commodification might be non-contingent. Consider the fact that coercion expresses contempt for its victim, but not on the basis of some merely conventional assignment of social meaning.)

which cooking as such characteristically produces.

In the case of cooperation, including cooperative exchange, each party to the exchange likewise acts as they do because of the characteristic further end of their action. For example, suppose two members of a household divide their labor so that, say, one person does the cooking and the other does the cleaning. Here, the cleaner will clean in the expectation and on the condition that the other will cook. But she will not clean for the sake of obtaining food. She will clean for the sake of maintaining the household, which is a characteristic and not a radically extrinsic end of cleaning.

The case of pure exchange, as we have already seen, is quite different. If one person cleans another's space merely so that the other will cook them dinner, the end of cleaning is no longer the maintenance of a household; it is food, nourishment. But cleaning as such does not aim at nourishment (unless it is, for instance, the cleaning of a cooking surface—but that is not the sort of way in which cleaning contributes to nourishment in a case of pure exchange). The end here is radically extrinsic to the means. And of course the same is true of the other party's action of cooking. Each party's contribution to a pure exchange, then, is motivated by a radically extrinsic incentive.

This would not be the case if exchange itself were a characteristic end of activities such as cooking and cleaning. And at least in a market society, one might be tempted to think that such labor, insofar as it is done for exchange, really is inherently *for* exchange. But that would be a mistake. It would be a version of the proposal, dealt with in Section 2.3, that coerced cooking is just an unusual form of self-defense. The suggestion now, in effect, would be that cooking for exchange is just an unusual form of cleaning. But it is

not the end of exchange, or the end of a clean kitchen, which identifies what concrete activity the cook is engaging in. What identifies her activity is only the end of nourishment. Only thus can we explain why the person who cooks for the sake of exchange (or for the sake of getting a clean kitchen in return) does not do the same thing as the person who moves furniture or gives a lecture or does anything else for the same purpose.

I am essentially making the same point here about useful activities which Aristotle makes in the *Politics* about useful things. He writes:

> Of everything which we possess there are two uses: both belong to the thing as such, but not in the same manner, for one is the proper, and the other the improper use of it. For example, a shoe is used for wear, and is used for exchange; both are uses of the shoe. He who gives a shoe in exchange for money or food to him who wants one, does indeed use the shoe as a shoe, but this is not its proper use, for a shoe is not made to be an object of barter.[28]

In suitable social conditions, anything that is useful can be done or produced for the sake of exchange. But even when it is exchanged, and even when it is produced for the sake of exchange, it does not follow that the object is *for* exchange. It is not the purpose of a shoe *qua* shoe to function as a medium of exchange. What makes it a *shoe*, as opposed to something else, is its use value, not its exchange value. By analogy, consider that anything useful, including a shoe, can in principle function as a way of making others envious, since its usefulness makes it a possible object of desire and, when conjoined with one person's lack and another's possession, a possible object of envy. It does not follow that a shoe, even in a very covetous society, has three purposes: use,

---

[28] Aristotle, *Politics*, , 1257a6–14.

exchange, and the incitement of envy.

Of course, like the characteristic ends of activities, the use value of things may be plural. A shoe may be protection for the feet, a fashion statement, a piece of athletic equipment, or a weapon for kicking protesters. But these are all purposes which it has in its concrete character as a shoe—and which will, for instance, determine specific features of its design—and not in its abstract character as a commodity.

8.2

Pure exchange, I have argued, is radically extrinsically motivated. It follows that it is unfree in fundamentally the same (distinctively interpersonal) way as coerced action, and morally problematic in fundamentally the same way as coercion. It would be useful to have a label for the common moral defect which I am arguing these two forms of interaction exemplify. One fitting description would be the familiar one of "using someone as a mere means." Both coercion and pure exchange are ways of exerting control over another's will in a way which impedes the other's exercise of her practical reason and so fails to respect her as a rational agent. As I will argue in the next section, they differ in morally salient ways. But they have this much in common.

Let us recall, in this connection, Grant's observation, quoted in Section 1.3, that incentives are essentially "instruments of government." Government may be consistent with respect for its subjects' status as agents if it secures their cooperation by giving them good reasons. But government by radically extrinsic incentives does not do so—whether the operative incentives happen to be sticks or carrots. Such government is heteronomy.

Another term which I think accurately describes both coercion and pure

exchange is "exploitation." On one reasonable understanding, to exploit someone is to take advantage of their vulnerability in a way which violates some duty of fairness or duty of protection toward that person.[29] But we owe it to each person to treat them in a way consistent with their free exercise of agency. In inducing them to do what we want using a radically extrinsic incentive, we take advantage of the fact that they have a certain need, the object of which we control and which therefore makes them vulnerable to us. And we violate the duty imposed by their right to freedom. So it is reasonable to say that we exploit them.

In the case of pure exchange (as opposed to coercion), the exploitation goes both ways. I think this kind of reciprocal taking advantage of one another's need is what Marx means by the provocative claim that in market exchange, "[t]he social relationship in which I stand to you, my work for your need, is […] a mere appearance and similarly our mutual completion is a mere appearance for which mutual plundering serves as a basis."[30] Pure exchange *looks* like cooperative exchange, but in reality it is a form of reciprocal exploitation.

Its reciprocal nature makes its exploitative nature more difficult to recognize. But the symmetry is often imperfect, and we can distinguish different degrees of exploitativeness just as we have already distinguished degrees of coerciveness: on the basis of the importance of the good being held hostage. When the parties to an exchange have unequal bargaining power, the exploitation which their exchange

---

[29] See, e.g., Robert E. Goodin, "Exploiting a Situation and Exploiting a Person," in *Modern Theories of Exploitation*, ed. Andrew Reeve (London: Sage, 1987).
[30] Karl Marx, "On James Mill," in *Karl Marx: Selected Writings*, ed. David McLellan (Oxford: Oxford University Press, 2000), 130.

involves will be asymmetric.

Consider the meeting between worker and capitalist on the labor market. The capitalist is looking to fill a job, and the worker is looking to get a job. But the loss which the capitalist faces if she fails to find someone for the job is likely to pale in comparison with the loss which the worker faces if he is unable to get a job. For this reason, the worker makes her contribution to the exchange more unfreely than the capitalist, and the capitalist exploits the worker to a greater degree than *vice versa*. If the inequality in bargaining power between two parties is severe, as it typically is in this case, it is reasonable to speak simply of the stronger party as exploiting the weaker—even though, strictly speaking, the weaker party also exploits the stronger.

By the same token, we may be *relatively* untroubled by pure exchanges which involve low stakes on both sides, since the degree of unfreedom involved in such exchanges is relatively low. Notwithstanding this difference of degree, though, the exchange remains a morally problematic *kind* of interaction on the basis of its motivational structure.

Incidentally, this gives us a more principled formulation of a criticism of so-called "noxious" markets which in its most common form can look quite arbitrary. Critics such as Michael Sandel and Debra Satz argue that market transactions entered into out of "dire" economic necessity are morally problematic because they are not fully voluntary.[31] But these philosophers are not prepared to reject markets in general, even

---

[31] Michael J. Sandel, "What Money Can't Buy: The Moral Limits of Markets," in *Tanner Lectures on Human Values*, ed. Grethe Peterson (Salt Lake City: University of Utah Press, 2000), Lecture I; Satz, *Why Some Things Should Not Be for Sale*, Chapter 4. See also

though market exchange is paradigmatically done from *some* degree of economic necessity. This raises the obvious question: what is so special about "dire" necessity, as opposed to merely middling or weak necessity, that it uniquely gives rise to unfreedom? It is a version of the same question I raised about the distinction between necessities and luxuries in Chapter III, Section 3.3.

My answer there also applies here. The difference is not, strictly speaking, a categorical difference. It is a difference of degree. It seems to me that we are prevented from seeing this only by an ideological commitment to the voluntariness of market transactions "in general." This commitment is difficult to square with the apparent involuntariness of exchange in conditions of dire need, so we are forced to cordon these off as uniquely special, "noxious" cases. But while they are indeed importantly different, the difference is one of degree.

8.3

This criticism of pure exchange may seem absurdly overhyped. What could be more obvious than that I do not wrong you or make you unfree in offering, for instance, to trade your chips for my pretzels at lunchtime, whereas I both wrong you and make you unfree in proposing to kill you unless you hand over your money?[32] Exchange *seems* to lack the unfreedom of coerced action and the wrongfulness of coercion. And the assumption that this appearance is correct constitutes common ground among virtually

---

Michael J. Sandel, *What Money Can't Buy: The Moral Limits of Markets* (New York: Farrar, Straus and Giroux, 2012).

[32] This is a fairly representative pair of examples from the literature. Pallikkathayil, "The Possibility of Choice," 1–2.

all of the theorists discussed in Chapter II.[33]

There are a few reasons to be more circumspect. First, it is easy, and extremely misleading, to conflate the various types of exchange I distinguished in Section 7. For instance, the lunchtime trade of chips for pretzels described above presumably takes place between schoolchildren or coworkers. Given some reasonable assumptions about the likely social relations between these individuals, this is quite likely to be a case of cooperative exchange rather than pure exchange. But as I have argued, it is only pure exchange which is essentially objectionable. As long as the exchange being entertained is a genuinely cooperative exchange, there is no problem.

Second, as I have already hinted, our intuitions about threats and offers are liable to be misled by a number of irrelevant factors. For example, we may tend to associate coercion with extremely high stakes—such as a mugging at gunpoint—and exchange with lower stakes—such as a trade of chips for pretzels. We may also tend to associate coercion with large inequalities of power or status, while associating exchange with equality. As I have argued, these factors make a genuine and important difference, albeit one of degree.

But the associations are strictly accidental. Coercion can involve low stakes, such as the threat to play a loud and obtrusive song which one knows one's "victim" dislikes; pure exchange can involve high stakes, such as access—or lack thereof—to the basic necessities of life. Coercion may be premised on equality, as in warfare between

---

[33] In addition to Pallikkathayil, see, e.g., Nozick, "Coercion," 459; Frankfurt, "Coercion and Moral Responsibility," 83; McGregor, "Bargaining Advantages and Coercion in the Market," 49n21; Wertheimer, *Coercion*, Chapters 12–13; O'Neill, "Between Consenting Adults," 273–274; and Ripstein, *Force and Freedom*, 36.

equally powerful parties; pure exchange may involve severe inequality, as between rich and poor. The more fairly we structure the comparison between coercion and pure exchange, the less obvious it seems that coerced action is normally involuntary whereas pure exchange is normally voluntary.

Third, living in a market society, we are constantly confronted, and constantly confront one another, with extrinsic incentives to action. We are accordingly under serious psychological pressure to accommodate ourselves to this condition. Our intuitions are liable to be distorted by a *status quo* bias. Moreover, our ability to imagine alternative forms of social relations—premised on more robust forms of reciprocity and cooperation—becomes impaired. In these conditions, we can only trust our common sense so far, and must turn to more systematic theoretical reflection.

But everything I have said up to this point amounts to a systematic argument that our common sense is indeed misled and misleading on this score. I have argued that the major extant theories of coercion fail to explain the unfreedom of coerced action or the wrongfulness of coercion, and that when their faults are corrected, we cannot but accept that what makes coercive interpersonal relations *pro tanto* wrong makes the interpersonal relations involved in pure exchange wrong in fundamentally the same way.

This is not to say that what is objectionable about these social relations might not be overridden in this or that particular context. Even if there is something *pro tanto* objectionable about entering into a pure exchange with someone, this does not immediately rule out the possibility that the interaction might be justified anyway, all things considered. For instance, in a social world hostile to genuine cooperation, neither

party might have a reasonable alternative there and then. The point is simply that there *is something* wrong with these interpersonal relations, and that it is the same thing which is most fundamentally wrong with coercive relations.

Nor is any of this to say that some instances of pure exchange can be morally innocent in certain *particular* respects. For instance, a pure exchange at "fair" prices lacks the special kind or degree of moral objectionableness and exploitativeness of an unfair exchange. But this does not imply that a "fair" exchange is morally innocent *tout court*, any more than the possibility of fighting a war according to certain rules of fairness implies that a state of war is as such unobjectionable.

Lastly, I am not suggesting that there is no moral difference at all between coercion and pure exchange in general. My point so far has simply been that whatever differences there may be, the difference is *not*: that coerced action is involuntary and pure exchange voluntary, or that coercion is objectionable and pure exchange morally innocent. Next, I turn to the question what the real difference might be.

9. THE MORAL DIFFERENCE BETWEEN COERCION AND PURE EXCHANGE

I have argued that both coercion and pure exchange are forms of exploitation. In this section, I argue that the primary moral difference between them concerns the scope of the exploited party's vulnerability.

9.1

Coercion and pure exchange are ways of taking advantage of another's vulnerability. Vulnerability is a relation which a person bears to one or more others with respect to

some object of need, i.e., some good. A person, *B*, is vulnerable to another, *A*, and is under *A*'s power, in the relevant sense, only if *B* has some need and *A* controls the object of that need. For instance, the victim of a mugger who says, "Your money or your life," is vulnerable to his mugger, and is under the mugger's power, because he needs to survive and the mugger has taken control of this good and now holds it hostage. The real moral difference between relevantly similar cases of coercion and pure exchange between *A* and *B* concerns the nature of *A*'s control of the object of *B*'s need.

Such control has what we might call a "positive" and a "negative" aspect: *A*'s (positive) ability to give *B* what *B* needs, and *A*'s (negative) ability to withhold it from *B*. And in either respect, *A*'s control can be either unilateral or only partial. *A*'s control is unilateral in the positive respect if *A*, by her own decision, can make it the case that *B* will get what he needs. *A*'s control is only partial in this respect if *A* requires the assistance of other agents or of fortuitous circumstances to make it the case that *B* will get what he needs. Similarly, *A*'s control is unilateral in the negative respect if *A*, by her decision alone, can make it the case that *B* will not receive the good. And *A*'s control is partial in this respect if *A* requires the assistance of other agents or of fortuitous circumstances to make it the case that *B* will not receive the good.

The fundamental difference between coercion and exchange, then, is this. If *A* coerces *B*, *A* exercises unilateral control over the object of *B*'s need in both respects. But if *A* merely engages in a pure exchange with *B*, *A* exercises unilateral control over the relevant good in the positive respect but only partial control in the negative respect: *A*

can unilaterally ensure that *B* gets the good, but not that *B* does not get the good.[34]

The mugger, for instance, arrogates to herself the power unilaterally to decide either that her victim will keep his life *or* that he will not, and she exercises both aspects of this power when she says, "Your money or your life." In particular, if the victim refuses to comply with her demand, she will unilaterally decide to end his life (I am assuming that she is prepared to follow through on her threat).

Contrast the case of pure exchange. For the purpose of illustration, let us suppose that a labor contract fits this model. Like the mugger, the capitalist who offers a worker money on the condition that he work for her exercises a measure of control over the object of the worker's need. However, this control is unilateral only in the positive aspect. The capitalist can unilaterally make it the case that the worker will have money. Her control is not unilateral in the negative respect. She cannot or will not unilaterally make it the case that the worker will have no money if he refuses the offer.

Of course, the capitalist's decision not to share her own wealth with the worker must make it in *some* way more unlikely or more costly for the worker to obtain this wealth—otherwise, her proposal would have no motivational efficacy at all. But the capitalist's ability to ensure that the worker ends up with no money is only partial. It depends on chance—such as the worker's not discovering a buried treasure or a rich aunt—and more importantly on the actions and omissions of other people—such as the

---

[34] This is not to say that *A*'s power to unilaterally make this the case is not itself dependent on certain social conditions. See, e.g., Richard Friedman, "Liberty Conceived as the Opposite of Slavery," in *Skepticism, Individuality, and Freedom: The Reluctant Liberalism of Richard Flathman*, ed. Bonnie Honig and David R. Mapel (Minneapolis: University of Minnesota Press, 2002).

failure of other capitalists to share their own wealth with the worker.

Though the stakes for the capitalist and the worker are, as I have emphasized, radically different, *structurally* the same things can be said about the worker's "offer" to the capitalist. The worker offers the capitalist his labor power to induce the capitalist to part with some fraction of her wealth. He can unilaterally make it the case that the capitalist has one unit of labor power at her disposal. But he has no unilateral power to make it the case that the capitalist lacks this unit of labor power if she refuses to pay him. That outcome would depend not just on the worker's own actions but on circumstances outside the worker's control—such as the availability of potential replacements for the worker—and on the actions of other people—such as those potential replacements' willingness to serve as replacements.

The main distinguishing feature of coercion *vis-à-vis* pure exchange, then, is the scope of the unilateral control which *A* exercises over *B*, and the scope of *B*'s vulnerability being exploited by *A*. In general, this makes the coercer's grip on the coercee's will more secure. Through her exercise of unilateral negative control, the coercer ensures that there are no, or very few, close possible worlds in which *B* does not have to comply with her demand in order to get the relevant good.[35] By contrast, the

---

[35] Gideon Yaffe makes a similar point—though I feel that he does not quite hit the nail on the head. He notes that "as a general rule, coercers don't merely produce, but also track, the compliance of their victims," being prepared "to threaten a more serious injury" should their initial threat fail to impress. Gideon Yaffe, "Indoctrination, Coercion and Freedom of the Will," *Philosophy and Phenomenological Research* 67, no. 2 (2003): 351. It may not be true of any given coercer, however, that she would be prepared to make a more serious threat than the one she actually makes. Nor does Yaffe's criterion help us distinguish coercion from exchange. The kind of "tracking" on which I am focusing, on the other hand, seems more essential to coercion, and also

parties to a pure exchange, in lacking or forgoing this comprehensive form of control, leave more to chance and to the possible intervention of other agents. Other things equal, there will be more possible worlds in which either chance or other people make it the case that *B* gets the relevant good without having to agree to the exchange.

This difference affects the significance of the wrong which *A* does to *B*, and explains the primary moral difference between coercion and pure exchange. Being prepared to ensure unilaterally that the coercee will not receive the relevant good unless she complies, the coercer displays an especially uncompromising commitment to subjecting the other's will to his own. In comparison, the parties to a pure exchange display something more like opportunism. They do not go out of their way to exacerbate another's vulnerability but merely take advantage of a preexisting vulnerability. Arguably, this is a less serious moral offense, other things equal.

9.2

We should not exaggerate the difference.

To some extent, my account of what distinguishes pure exchange from coercion aligns with a classical liberal argument in defense of the market. Theorists from Adam Smith through Milton Friedman to Elizabeth Anderson and Debra Satz have pointed out that individuals transacting on the market are less vulnerable to one another as individuals than slaves and serfs are to their masters and lords. In Smith's words:

> Each tradesman or artificer derives his subsistence from the employment, not of one, but of a hundred or a thousand different customers. Though in

---

distinguishes it from exchange: namely, that a serious coercer is prepared to go out of her way to make sure that the undesirable consequence is inflicted on her victim in all or very many of the possible worlds in which the latter refuses to comply.

> some measure obliged to them all, therefore, he is not absolutely dependent on any one of them.[36]

These thinkers have tended to infer that market exchange as such is completely morally innocent and voluntary, and in particular that the direct interpersonal relations between transacting parties are unproblematic.

This impression may be strengthened by a distinctive feature of pure exchange which concerns the assignment of moral responsibility for each party's vulnerability to the other. As I have mentioned, the success of *A*'s attempted inducement of *B*'s action in a pure exchange typically depends on the failure of other people to intervene. For instance, the capitalist's attempt to induce the worker to work for him depends for its success on the fact that the other owners of capital who are in a position to share it with the worker refuse to do so. As a result, in cases of pure exchange, agents other than *A* will typically bear partial responsibility for *B*'s predicament. More confusingly still, as Julius points out, it may be the case in principle that no identifiable person at all, including *A*, is responsible for *B*'s vulnerability.[37] No identifiable capitalist, for instance, may have an independent moral obligation to share their capital *with B in particular*. If *A* and *B* had never interacted and exchanged, it might be the case that *A* in particular would not have violated any personal moral obligation to *B* in particular.

It is easy, against this background, to lose one's grip on what is wrong with the

---

[36] Smith, *The Wealth of Nations*, 438. Compare Friedman and Friedman, *Capitalism and Freedom*, 14–15; Satz, *Why Some Things Should Not Be for Sale*, 42; and Elizabeth Anderson, *Private Government: How Employers Rule Our Lives (and Why We Don't Talk About It)* (Princeton: Princeton University Press, 2017), Chapter 1.
[37] A. J. Julius, "A Form of Unfreedom," (Manuscript): §11.

interpersonal relation between *A* and *B* in particular. For those with an essentially interpersonal conception of freedom, this will in turn tend to make it hard to see that *B* could really be unfree. But there is something wrong with the interpersonal relation. By using the object of *B*'s need as a radically extrinsic incentive to control *B*'s will, *A* takes advantage of *B*'s vulnerability in a way which makes *B*'s action unfree. Irrespective of the question who if anyone is responsible for *B*'s vulnerability in the first place, this is no way to relate to another person.

Now, since a pure exchange is symmetrical as far as this motivational structure is concerned, the same will be true in the other direction. So *A*'s exploitation of *B*, by virtue of *B*'s exploitation of *A*, is itself chosen unfreely. This makes it even harder to settle the question of *A*'s (and *B*'s) moral responsibility or culpability for the wrong. But the issue of responsibility is a red herring as far as the real problem with the relation between *A* and *B* is concerned. The possibility of pure exchange shows that it is possible unfreely to wrong another person, and unfreely to make them unfree.

We are finite, interdependent creatures. Inevitably, therefore, we will have needs and on occasion we will have control over objects of one another's needs. Some ways of exercising this control are consistent with relating to one another as free agents; others are not. A minimal condition of treating others with the requisite respect is that one attempt to get others to do something only in ways formally consistent with their doing it for its own sake, on the basis of their own judgment of its merits. Both coercion and pure exchange fail to meet this standard. To the extent that our social relations widely take the form of either coercion or pure exchange, then, we do not live in a society of free people.

CHAPTER V. TOIL

The topic of this chapter is "material" unfreedom: the unfreedom which an activity can have by virtue of its content. Even though this dimension of freedom and unfreedom is not concerned directly with the character of interpersonal relations, it can still give rise to second-personal claims. I illustrate this using the example of toil—labor which is intrinsically unchoiceworthy by virtue of its content. I describe two contemporary forms of toil and explain what makes them unchoiceworthy (Section 1). I then defend three imperatives of freedom in relation to toil: a duty of elimination and transformation, a duty of reduction, and a duty of redistribution. First, I argue that we have a collective duty to eliminate pointless toil and transform useful toil into meaningful work where possible (Section 2). I suggest that this is likely to require substantial changes to the basic structure of our society (Sections 3–4). I then take a brief detour to relate my position to various theoretical strands of the "anti-work" tradition (Section 5). Finally, I consider the possibility that some toil may be unavoidable at present. I argue that if there is such toil, we have a collective duty to reduce it in the long run and to distribute it justly in the meantime (Section 6).

1. TWO KINDS OF TOIL

The idea of freedom as activity for its own sake has ethical implications that concern not just the motivational structure of our interpersonal transactions but also the content of our activities. In particular, it has implications for what *labor* should look like if it is to

be free. In this section, I argue that labor which the worker does not think serves a good purpose or which is extremely specialized is unfree.

1.1

In Chapter III, I argued that the perceived intrinsic value of an instrumental action, if it has any, must be grounded in the perceived instrumental value that gives the action its identity. This claim has the obvious consequence that if a productive activity is to be done for its own sake, its product must be perceived to have some value. The labor, in other words, must be perceived to be genuinely useful.

This seemingly trivial condition would seem to be violated by the many contemporary occupations which David Graeber has memorably dubbed "bullshit jobs." He defines these as

> jobs that are primarily or entirely made up of tasks that the person doing that job considers to be pointless, unnecessary, or even pernicious. Above all, these are jobs that the holders themselves feel should not exist.[1]

Some examples of bullshit jobs which Graeber suggests are "HR consultants, communications coordinators, PR researchers, financial strategists, [and] corporate lawyers."[2] No doubt one can argue about the details of any proposed list of "bullshit jobs," and no doubt there will always be exceptions within any given occupational category. It is significant, however, that many workers in industrialized countries appear to think of their own work as essentially pointless or harmful. Graeber mentions recent surveys finding that 37% and 40% of British and Dutch workers, respectively, do

---

[1] David Graeber, *Bullshit Jobs: A Theory* (New York: Simon & Schuster, 2018), 29.
[2] Ibid., 15.

not think that their jobs "make a meaningful contribution to the world."[3]

In general, work whose product the worker does not consider valuable cannot be done for its own sake. Bullshit jobs are instrumental activities (albeit defective ones) and as such are defined by their characteristic product. If they are not valued under the description of producing this product, then they are not valued intrinsically and not done for their own sake. It follows, given my account of freedom, that bullshit jobs are done unfreely, under the guise of *mere* necessity.

Of course, workers generally do find *some* instrumental value in bullshit jobs; after all, they *choose* to work in these occupations. But the value such work has for the worker is bound to be some radically extrinsic end (as defined in Chapter IV) such as paying the bills. Indeed, the fact that it is motivated in some such radically extrinsic way is what makes it possible for such labor to exist at all. Normally, one would expect an agent who saw no value in the product of a given productive activity not to choose to engage in that activity. What explains why people choose bullshit jobs in spite of their perceived pointlessness is not the perceived value of the (characteristic) product of the activity but rather some end radically extrinsic to the activity.

The fact that radically extrinsic motivation must be involved in this way shows that bullshit jobs, though they are materially unfree (by virtue of their concrete worthless content), are actually a symptom of formal, interpersonal unfreedom. The worker does the work only because someone else—presumably her employer—artificially connects it with ends which the worker does value. No such labor could exist

---

[3] Ibid., 19.

in a society in which people related to one another as free agents. People who are their own masters do not choose to do what they consider pointless.

1.2

For work to be done for its own sake and freely, however, it is not sufficient that its product be perceived as good. The worker must also have this further good in view as she works. This condition, the content of which will become clearer below, is typically violated by jobs that fall under the heading of "routine labor."

Routine labor is labor which consists only of relatively simple, prescribed operations, involving little scope for the exercise of skill or deliberation on the part of the individual worker. To show how my account of freedom gives us a diagnosis of what is wrong with work of this kind, I will discuss two examples I have already had occasion to mention: Phil, the American spot-welder, and T, the British call center specialist.

Phil's work, to reiterate, essentially consists in standing in one spot and pushing a button about 12,000 times a day.[4] T, who helps troubleshoot customer's problems with phone and broadband connections on the phone, has a similarly restrictive job description:

> T is supposed to answer and deal with whatever arises after a call is finished in a certain time, he has to direct the problem to the person who can solve it on the first call; he must collect details like mobile numbers, email addresses and assent to a text message customer satisfaction survey. "You have customers who don't want to listen but you've got managers breathing down your neck, saying you've got to let them speak to you like crap basically." A noticeboard at the entrance divides ordinary phrases into two columns: 'I'm not sure', 'usually', 'obviously' and 'maybe' are

[4] Terkel, *Working*, 159.

> "below the line" and ought to be avoided. "Above the line" terms include: 'lovely', 'excellent' and 'what can I do'?[5]

What is it like to do such work? As David Frayne and others have observed, one of its characteristic phenomenological features seems to be something like the opposite of a "flow state."[6] Rather than becoming absorbed in one's work and losing one's sense of time independent of the rhythms of the work itself, one remains at an arm's length from the work, seeking distractions and counting down the hours and minutes until the next break or the end of the working day. Phil emphasizes the need for daydreaming:

> You pretty much stay to yourself. You get involved with yourself. You dream, you think of things you've done. I drift back continuously to when I was a kid and what me and my brothers did. The things you love most are the things you drift back into. … Repetition is such that if you were to think about the job itself, you'd slowly go out of your mind.[7]

Pervading T's description of his work, similarly, is the constant need to get away from it:

> "To be honest, this place takes so much out of me," he says. "[…] unless I've got a couple of days off together, I do feel I've lost a large part of myself working here. I just had two weeks off through sick and holiday, and it was painful coming back. I really, really had to make myself come in. Because I just didn't want to." […] He gets up at 6 a.m. most days and walks the two and a half miles to work listening to music: "The music is to stop me thinking about the fact that I'm heading here." […] He must be logged in and ready to take calls at 8 a.m. Somewhere between 9.15 and 10.45 a.m. he'll have a break for fifteen minutes […]. He used to have a coffee ever hour "because it was an excuse to get away from my desk"; now he has cut down to three or four a day. […] If there's a particularly nasty call, he'll go out and smoke, hoping someone will be out there:

---

[5] Biggs, *All Day Long*, 84, quotation marks modified for clarity.
[6] David Frayne, *The Refusal of Work: Rethinking Post-Work Theory and Practice* (London: Zed Books, 2015), 12. See Mihaly Csikszentmihaly, *Flow: The Psychology of Optimal Experience* (New York: Harper and Row, 1990).
[7] Terkel, *Working*, 160.

> "That's really the only way to survive it when you're getting calls like that. You need to vent about them."[8]

This cannot be what it is to do work for its own sake. There are many interconnected problems with it, but the one I will focus on here concerns the relation—or lack thereof—of the individual worker to the (further) good which their work serves. Neither Phil's nor T's work is a bullshit job. Phil contributes to the production of cars and T to the maintenance of phone and broadband connections. Yet both Phil's and T's jobs are so specialized—the more general productive processes in which they participate so compartmentalized—that as individual workers they become disconnected from these social contributions.

In particular, they are not encouraged (indeed they are forbidden) to engage in individual deliberation or exercise a skill with a view to this further good. That is, they are not given an opportunity to allow their own apprehension of this further good to have any practical significance, to guide their concrete activity in any meaningful way. No matter how much a worker in Phil's position cares about making cars and understands what it is to make a good car, this concern and knowledge are essentially irrelevant to the job of repeatedly pressing a button. Understandably, Phil expresses a desire for work which would at least come closer to enabling him to bring the end of making (say) a good car and the knowledge of how to do it to bear on his actual practice:

> Funny things is, I don't mind working at body construction. To a great degree, I enjoy it. I love using my hands—more than I do my mind. I love to be able to put things together and see something in the long run. I'll be the first to admit I've got the easiest job on the line. But I'm against this

---

[8] Biggs, *All Day Long*, 83–85.

> thing where I'm being held back. I'll work like a dog until I get what I want. The job I really want is utility.
>
> It's where I stand and say I can do any job in this department, and nobody has to worry about me. As it is now, out of say, sixty jobs, I can do almost half of 'em. I want to get away from standing in one spot. Utility can do a different job every day. Instead of working right there for eight hours I could work over there for eight, I could work the other place for eight. Every day it would change.[9]

T's scope for deliberation at work, while perhaps slightly greater than Phil's, is also severely limited: for instance, by the strict time benchmarks, the checklist of required tasks, and the list of officially encouraged and discouraged phrases to use in conversations with customers. As Biggs notes, writing about T: "The emotional labour of a call centre isn't just learning to withstand a customer's impatience but also learning to speak in a voice other than your own."[10] More generally, the extreme isolation of T's responsibilities from the broader productive process to which they belong—the process of providing people with functioning phone and broadband connections—makes it difficult or impossible to conceive of what he is doing as guided by the good which this process as a whole is meant to serve:

> T doesn't feel satisfied at close of day: "You've got no vision of what will happen to a call once you put the phone down. It goes off somewhere else, someone else deals with it and whether they get things right, get things wrong, get it fixed, you don't know."[11]

Now, I submit that work done in this way—where one's activity is so rote and prescribed that it cannot meaningfully be guided by one's own apprehension of the good which the activity is meant to produce—cannot be done for its own sake. For

---

[9] Terkel, *Working*, 162.
[10] Biggs, *All Day Long*, 84.
[11] Ibid., 85.

when one's activity becomes practically detached from its further end in this way, it ceases to be, from the worker's own point of view, a way of helping to make a car or helping to maintain someone's broadband connection, and becomes instead a *distinct* activity identified by some more impoverished description such as "pushing a button" or "going through a checklist with a stranger on the phone." And when done under some such impoverished description, the activity is no longer done under the description which makes it instrumentally good. But then it cannot be done for the sake of any intrinsic value which attaches to it on account of this instrumental value. So it cannot be done for its own sake.

An extreme case of this phenomenon would be that of a line worker in a factory who does not even know what it is that she is helping to make, perhaps because she does not know the local language.[12] It is difficult to see how such a worker could be doing what she was doing for the sake of making the product, and, by the same token, how she could be engaged in making that product for its own sake. But though Phil and T of course happen to have the relevant knowledge, their situation is, in practical terms, more or less equivalent to that of a worker who lacks that knowledge.

What gives rise to this situation is an extreme form of the division of labor. While some division of labor is necessary in any large society, since presumably no individual could perform all the necessary types of work even over a whole lifetime, this must be distinguished from the "detailed" division of labor within a given occupation. This division separates the tasks of planning from the tasks of execution and allocates these

---

[12] Graeber claims that there are "many" such cases in the U.S. *Bullshit Jobs*, 37.

tasks to distinct people. The social division of labor among different occupations in principle keeps intact the unity of practical reasoning on the one hand with the activity which that reasoning governs on the other. The detailed division of labor destroys that unity. Being cut off from the planning aspects of her work, the detail worker no longer has the opportunity to govern her work by her own practical reasoning on the basis of her apprehension of the purpose of the work. But in this way, her activity ceases to be done by her for the sake of this end, and consequently also ceases to be intrinsically choiceworthy as a certain kind of activity. For that reason, it can no longer be done freely.

While this dissertation is not an empirical study of labor, it certainly seems likely that bullshit jobs and routine labor between them make up a large proportion of the work people currently do, both in the U.S. and around the world. If this is true, then it cannot be truly said that we live in a society of free people, quite apart from the exploitative and coercive interpersonal relations which I analyzed in the previous chapter. Liberation would require transforming not just our relations to one another but also the structure of concrete work processes. In particular, it would require doing away with work which manifestly serves no good purpose and with the extreme specialization of labor, both of which are inconsistent with the freedom of the individual worker.

2. Eliminating and transforming toil

If individuals have a right to freedom and freedom is what I have said it is, then individuals have a right not merely not to be coerced or exploited but also not to have to

perform intrinsically unchoiceworthy labor such as bullshit jobs and routine work, to the extent that this can be avoided. To what more specific claims does this right give rise, and who can address these claims to whom?

Insofar as the work in question is paid employment, workers in bullshit and routine jobs can make certain claims against their employers. These employers control certain goods which their workers need, and they choose to make their workers' access to these goods conditional on the performance of work which cannot be done freely. As I argued in the previous chapters, such employers already wrong their workers by using the objects of the workers' need as radically extrinsic inducements. Employers of bullshit and routine labor, however, commit an additional wrong by using their power over workers to extract this particular *kind* of labor—labor which is in its very nature unfree. It adds another insult to the workers' dignity as agents, and workers have a direct claim against their employers that the latter not use their power over them in this way.

That said, of course employers' choices to extract bullshit and routine labor from workers do not occur in a vacuum. Given how widespread these kinds of toil appear to be, one can surmise that employers are themselves responding—in a systematic way—to a perceived necessity independent of them which arises from some aspect of their environment. While being constrained or compelled in some such way may or may not make them less morally culpable, it certainly means that the employers are unlikely to change their ways of their own accord, and so unlikely to be *effective* addressees of the claims which they violate.

It would seem, then, that any realistic effort to overcome the unfreedom of toil

must target the general social conditions which originally bring it into being. Accordingly, looking beyond the one-to-one interpersonal relation between worker and employer, workers in bullshit and routine jobs also have claims against their fellow citizens (including against one another) collectively to do what they can to transform these social conditions. As with all positive duties, including duties to prevent others' negative rights violations, it is difficult to say in general who exactly has a responsibility to do how much. In the absence of an effective central coordinating agent, much will presumably depend on who is able to do what.

Putting these issues aside, however—for they are simply too large for me to tackle here—we can say at least the following. First, all of us together have a collective duty to transform our social conditions so as to eliminate both bullshit labor and the detailed division of labor. Second, owing to this collective duty, each of us individually has a duty, in the absence of some reasonable excuse, to make some contribution to this goal. But what does this goal amount to? What would the requisite social transformation look like?

3. Class conflict as a cause of toil

This is not the place for a full study of the social causes of bullshit jobs and routine labor. However, in order to begin to show what the abstract imperative to eliminate or transform toil might look like more concretely, I will identify two pressures stemming from the basic structure of our society which I hypothesize tend to produce such work. The first source of pressure, which is the topic of this section, is the division between the working class and the capitalist class, between those who labor and those who employ

labor.

3.1

When workers work not for the sake of the product of their work—for the sake of making (good) cars, say—but for the sake of making a living, and when capitalists, for their part, employ workers not for the sake of the product either but for the sake of making a profit, then their interests can be expected to diverge in important respects. The many specific manifestations of this divergence of interests hardly need rehearsing, as they are visible all the time in disputes between management and organized labor concerning pay, benefits, hours, job descriptions and titles, job security, workplace safety, workplace surveillance, grievance procedures, and so on.

To see the distinctiveness of such conflict, consider what it would be like if there were no separation between workers on the one hand and capitalists and top managers on the other. Those making decisions about working conditions would be the workers themselves. In this situation, conflicts about these matters could not exist in the same form. Presumably individual workers would still disagree amongst themselves on this or that issue, and various factions might be expected to form along various boundaries of disagreement, but conflicts would not consistently pitch one group of decision-makers with a consistent set of interests against another group of decision-takers with a consistently opposed set of interests. The single most entrenched and consequential source of conflict in the contemporary workplace could not exist in these circumstances.

One immediate relevant consequence of class conflict in our society is that capitalists have a need to employ a substantial amount of labor for the specific purpose of disciplining their workforce and otherwise preventing class conflict from becoming

excessively disruptive. At least some of these special occupations are likely to make up an important category of bullshit jobs. One plausible example is the work of corporate lawyers and corporate lobbyists, or at least a significant portion thereof. If and when the primary occupation of lawyers and lobbyists becomes the defense and facilitation of their employers' exploitation of workers in the face of organized labor, political opposition, and other forms of resistance, they perform work which they might very reasonably think makes no valuable social contribution, though at least it pays well. If and when they do think so, their jobs are bullshit jobs. But this portion of the practice of corporate law and lobbying would not exist but for the existence of class conflict.

Class conflict also arguably gives rise to a portion of the bullshit occupations which Graeber calls "box tickers": "employees who exist only or primarily to allow an organization to be able to claim it is doing something that, in fact, it is not doing."[13] A paradigm case of such work would be the administration of a grievance procedure which is specifically designed not to address the root causes of the problems which give rise to the grievances. Now, capitalists have an interest in defusing resentment among their workforce and among the labor movement more generally, but without making substantive changes to the power structures which fuel this resentment. They thus have an incentive to create box ticking exercises of just this sort, and hence box ticking jobs. Some of these jobs may be in the public sector, within one or another special "commission," "board," or "committee." If and when such exercises are indeed designed merely to give the appearance of solving a problem, and if and when those

[13] Ibid., 81.

who staff these bodies recognize this fact, then their labor is a plausible candidate for bullshit work—one which, again, is likely to be bloated by class conflict.

3.2

In addition to giving rise to bullshit jobs, the class division between capitalists and workers also generates pressure toward the routinization of labor. A key locus of class contestation is the shape of the concrete work process itself. Particularly important here is the rate of production. In general, it is in the interest of capitalists for the work process to be as productive as possible: to maximize output for a given quantity of inputs, including labor time, no matter the human cost to the worker in terms of fatigue, pain, illness and injury, degradation, and loss of meaning—these human costs being relevant only indirectly and contingently, insofar as they impact productivity. Workers, on the other hand, obviously have an interest in limiting the rate of production in order to protect themselves. For a stark illustration of this conflict of interests, one need only look at the working conditions in Amazon warehouses, where workers are compelled to move merchandise at a near-impossible and unrelenting speed. According to reporting based on internal data from 23 of Amazon's 110 U.S. "fulfillment centers," the company recorded "9.6 serious injuries per 100 full-time workers in 2018."[14]

In light of this conflict of interests, capitalists have an incentive to find ways of controlling the work process as directly and comprehensively as possible, leaving as

---

[14] Will Evans, "Ruthless Quotas at Amazon Are Maiming Employees," *The Atlantic* 2019. On the other hand, Jeff Bezos recently signed a statement by the Business Roundtable committing to "investing in our employees" and to fostering "dignity and respect." Business Roundtable, "Statement on the Purpose of a Corporation" (2019).

little as possible to the decision-making of a resistant workforce. The routinization of labor is a singularly effective means to such control. Prior to routinization, management's control is limited by the fact that the work process depends on knowledge which is concentrated in the individual skilled craftsperson. By reducing the individual worker's labor to a series of simple, predetermined tasks which can be easily prescribed and measured and which leave the individual worker with no scope for meaningful deliberation, management ceases to be dependent on individual workers' knowledge and decision-making and effectively takes control over the work process out of their hands. In particular, management gains comprehensive and minute control over the rate of production.

This capitalist rationale for routinization is not merely a just-so story. Braverman argues convincingly that the extension of management's direct control over the work process motivated the routinization of labor as a matter of actual history.[15] He shows, for instance, that it was quite nakedly the goal of Frederick Taylor's pioneering work in scientific management. In *The Principles of Scientific Management*, Taylor observes:

> Now, in the best of the ordinary types of management, the managers recognize frankly that … the workmen, included in the twenty or thirty trades, who are under them, possess this mass of traditional knowledge, a large part of which is not in the possession of management. […] They recognize the task before them as that of inducing each workman to use his best endeavors, his hardest work, all his traditional knowledge, his skill, his ingenuity, and his good-will—in a word, his "initiative," so as to yield the largest possible return to his employer.[16]

---

[15] Harry Braverman, *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*, 25th Anniversary ed. (New York: Monthly Review Press, 1998), especially Chapters III–IV.

[16] Frederick W. Taylor, *Principles of Scientific Management*, in *Scientific Management* (New York and London: 1947), 32, quoted in Braverman, *Labor and Monopoly Capital*, 70.

In effect, Taylor concluded, this type of shop is "really run by the workmen and not by the bosses. The workmen together had carefully planned just how fast each job should be done."[17] Taylor proposed to overcome this limitation:

> Perhaps the most prominent single element in modern scientific management is the task idea. The work of every workman is fully planned out by the management at least one day in advance, and each man receives in most cases complete written instructions, describing in detail the task which he is to accomplish, as well as the means to be used in doing the work…. This task specifies not only what is to be done, but how it is to be done and the exact time allowed for doing it.[18]

A century later, the routinization of labor continues apace even in comparatively new, initially "skilled" occupations such as programming. And as long as the division between capitalists and workers endures, there is some reason to expect the trend to continue. For in these conditions, those who decide what work looks like—constrained only partially and variably by a process of negotiation and political struggle—have an interest in taking decision-making in work away from those who do the work.

4. THE MARKET ECONOMY AS A CAUSE OF TOIL

In this section, I identify a second feature of the basic structure of our society likely to give rise to avoidable toil: the market economy.

4.1

The diagnosis of the social causes of toil in the previous section is incomplete, for it

---

[17] Taylor, *Principles of Scientific Management*, 48–49, quoted in Braverman, *Labor and Monopoly Capital*, 70.
[18] Taylor, *Principles of Scientific Management*, 63, quoted in Braverman, *Labor and Monopoly Capital*, 82.

presupposes and does not explain why capitalists' fundamental interests diverge from those of workers in the first place. For instance, how is it that capitalists come to have an interest in the unconstrained expansion of the rate of production in the first place? Coming at great human cost, even if that cost is borne by other people, the pursuit of such a goal by human beings would in any other context strike us as unusual—indeed, positively sociopathic—and calls for explanation. What explains this and other interests of capitalists is the market economy.

The engine which drives production in a market economy and determines its form and content is competition, which forces capitalists to subordinate their other ends to the accumulation of capital. For example, assuming a perfectly competitive market and holding other things equal, a company whose employees work at a lower rate can expect to find its competitors able to charge lower prices, capture a bigger market share, and eventually drive it out of business. To avoid this fate, it must keep pace with its competitors' productivity. Similar dynamics apply to the other loci of class struggle. A company that pays its employees more than its competitors will have higher costs, other things equal; its competitors will be able to charge lower prices and eventually drive it out of business (again, other things equal). So it is the market which inflexibly gives rise to the peculiar interests which pit capitalists against workers. As Marx puts it, "the capitalist is just as enslaved by the relationships of capitalism as is his opposite pole, the worker, albeit in a quite different manner."[19]

The dynamics of competition give rise to the historical economic advantages of

---

[19] Karl Marx, *Capital: A Critique of Political Economy*, trans. Ben Fowkes, vol. 1 (New York: Vintage, 1977), 990.

the market, such as the massive increase in productivity since the emergence of capitalism. But in light of the previous discussion, we can see that in a context of class division, the same dynamics will tend to give rise to bullshit jobs and routine labor. Of course, the divergence of interests between "employers" of other people's labor and those whose labor they "employ" exists in non-market contexts too: for instance, in ancient slaveholding societies between slaveholder and slave, or in feudal societies between lord and serf. But due to the pressure of competition, the "employer's" demands on the "employed" become distinctively boundless in the context of a market economy.

4.2

In fact, the market will also have a tendency to generate toil independently of the class dynamic between capitalists and workers. First, one effect of the competitive imperative to accumulate capital is likely to be the proliferation of work whose purpose is not to satisfy a pre-existing need. We can see this by examining an argument of G.A. Cohen's, following Marx. Cohen notes that increases in productivity generally give producers the opportunity to increase leisure—which he defines as "freedom from unappealing activity," opposed to what he calls toil, which is "activity insofar as it is unappealing"—or to increase output.[20] But "capitalism," he argues, "inherently tends to produce just one of the options, output expansion, since the other, toil reduction, threatens a sacrifice of the profit associated with increased output and sales, and hence a loss of competitive

[20] G. A. Cohen, *Karl Marx's Theory of History: A Defense*, Expanded ed. (Princeton: Princeton University Press, 2000), 303–304.

strength."[21] Specifically, what is responsible for this tendency is not the class division between capitalists and workers but the imperative to accumulate capital—which has its source in competition on the market. Thus, one could expect worker-run firms competing against one another in a free market to find themselves consistently compelled to make the same trade-off.[22] As Marx puts it, in such firms "the opposition between capital and labour is abolished […] at first only in the form that the workers in association become their own capitalist."[23]

But if this is so, then producers will tend to have an incentive, in any market economy, to produce more *regardless* of whether there is a pre-existing need to satisfy. This can generally be expected to generate more work than is really necessary to satisfy people's needs. When workers recognize that what they are producing is not being produced in order to satisfy genuine needs, then their jobs will qualify as bullshit jobs.

Relatedly, the task of actually selling all of the indefinitely increasing output being produced itself requires work: the work of deliberately producing *new* needs which the products being produced can satisfy. This is the work of marketing—surely a prime sphere of bullshit jobs which are widely agreed to make no valuable social contribution. But competition can bring such work into being, in something akin to an arms race.[24] Given the possibility that one's competitors might conduct marketing

---

[21] Ibid., 304.
[22] Ibid., 314.
[23] Marx, *Capital*, 3, 571. Marx says "at first" because he thinks worker-run factories are bound to undergo further transformation, since they are the incipient emergence of a genuinely non-capitalist form of social organization. But that is beyond the scope of my present discussion.
[24] Graeber specifically compares marketers to armies, which any given country only needs because other countries have them. Graeber, *Bullshit Jobs*, 71.

campaigns, one risks losing out if one does not do so oneself. Even if everyone would in fact rather spend no resources at all on marketing (and people would rather not be subjected to it), this outcome is not a stable equilibrium of a competitive market. The same sort of dynamic, one might add, applies to large portions of corporate law, corporate lobbying, and PR work.[25]

4·3

In addition to generating bullshit jobs, the market tends to cause the routinization of labor, for such routinization can be more productive, in the sense that a greater output can be generated from given inputs, including labor time. Braverman distinguishes at least two ways in which it is more productive. First and most obviously, routinization can save the labor time which a "complete" craftsperson would take to switch from one operation in the productive process to another. In a given amount of time, therefore, more output can be produced. Of course it is an ethical question whether this increase in productivity is worth the impoverishment of the content of the labor, but the effect itself is a technical fact; in particular, it does not depend on the separation between capitalists and workers. A worker-run firm in a market economy would therefore face the same trade-off and be subject to the same pressure to routinize its labor with a view to increasing productivity.

Second, routinization is also efficient in the sense that it cheapens labor power. In general, at least given a sufficiently large scale of production, it costs more to hire a single "complete" craftsperson than it costs to hire whatever number of routine laborers

---

[25] Ibid.

it takes to perform the same amount of labor. In the words of the nineteenth-century political economist Charles Babbage:

> the master manufacturer, by dividing the work to be executed into different processes, each requiring different degrees of skill or of force, can purchase exactly that precise quantity of both which is necessary for each process; whereas, if the whole work were executed by one workman, that person must possess sufficient skill to perform the most difficult, and sufficient strength to execute the most laborious, of the operations into which the art is divided.[26]

Now, Braverman suggests that this second, cheapening effect of routinization is specific to "a society based upon the purchase and sale of labor power."[27] But the effect is certainly not specific to a society in which capitalists (the purchasers of labor-power) and workers (its sellers) are divided against each other as *separate* groups of people. To be sure, in a genuine class society, the effect can be expected to be stronger, owing to historical patterns of inequality. The craftsperson is likely to have disproportionately more bargaining power than those most likely to perform the more degraded labor—historically, for instance, women, children, and immigrants. Yet the same *type* of cheapening effect of routinization would be seen in a market economy of worker-run co-operatives in which there was no wage bargaining between employer and employee, assuming only that each of these co-operatives was responsible in one way or another for the costs of training its own workforce.

For at least these two reasons, then, the competitive pressure of the market will tend to compel producers to routinize labor. For this reason, it seems to me that the sort

---

[26] Charles Babbage, *On the Economy of Machinery and Manufactures* (London, 1832; repr., New York, 1963), 175–176, quoted in Braverman, *Labor and Monopoly Capital*, 55.
[27] *Labor and Monopoly Capital*, 55.

of workplace reforms defended by other champions of meaningful work such as Adina Schwartz would likely be insufficient. Schwartz argues that in order to make work compatible with the autonomy of individual workers, "jobs must be democratically redesigned, tasks must be shared out in a way that abolishes the distinction between those who decide and those who execute others' decisions."[28] This seems right as far as it goes, but it is clear that she imagines democratically run firms competing against one another in a market economy.

If the arguments I have sketched are on the right track, this solution would fail to address the root cause of the problem. If the market generally compels efficiency, then even workers running their own fully democratic firms—with no division between those who decide and those who execute others' decisions—competing against other similarly structured firms will find themselves constantly subject to a compulsion to routinize their own labor—to be "their own capitalists." While workers might formally be making their own decisions at the level of the operation of the firm, the concrete content of their labor—in particular, the scope for deliberation and the exercise of skill within it—might ultimately be little affected.

The argument against the hierarchical organization of the workplace (what Marx calls the "despotism" of the firm), then, would also seem to be arguments against the market organization of social production (what he calls the "anarchy of the social division of labor").[29] Of course, in the form in which I have presented them here, these arguments remain far from conclusive still. My aim has merely been to suggest some

---

[28] Adina Schwartz, "Meaningful Work," *Ethics* 92, no. 4 (1982): 640.
[29] Marx, *Capital*, 1, 477.

reasons for thinking that the market economy tends to generate certain *pressures*—pressures which could explain a state of affairs that cries out for explanation, namely the apparent prevalence of toil in a supposedly rationally ordered and democratic society, at a time when our collective technical capability to liberate people from this unfreedom ought to be unprecedented.

5. FREE TIME AND ANTI-WORK THEORY

I have defended a collective duty to eliminate certain kinds of labor, on the grounds of their unfreedom. This goal must be distinguished, however, from that of increasing "free time," if this is understood essentially as time *outside* of work. In order to make this clear, I will now compare my position with a diverse range of positions within what may be called an anti-work tradition. Many of these theories seem to me in some way to miss or misstate the most important point of the social critique of toil.

5.1

Some philosophers argue, in effect, that labor *as such*—or at least "necessary" labor—amounts to toil and therefore cannot be done freely.

As I mentioned in the Introduction, Aristotle suggests that labor aiming at the sustenance and reproduction of life is really fit only for slaves and animals, and that the highest good, by contrast, consists in contemplation, since this activity is done purely for its own sake.[30] In the twentieth century, Arendt follows Aristotle in defending the

[30] Aristotle, *Politics*, 1254b25; *Nicomachean Ethics*, 1177b30ff.

comprehensive unfreedom of both what she calls "labor"—the activities of sustaining and reproducing life itself, such as farming, cooking, eating, and procreating—and what she calls "work"—the activities of producing the material environment in which we live and which we use—such as the making of tools, houses, clothes, and cars. All of these activities, she argues, are done under the guise of necessity and therefore not freely chosen; they cannot express our own spontaneous agency but only our subjection to external forces and limitations.[31]

More surprisingly, one finds fairly direct echoes of the same sort of idea in at least one place in Marx's work. In a beautiful but flawed passage of Volume 3 of *Capital*, Marx writes:

> The realm of freedom really begins only where labour determined by necessity and external expediency ends; it lies by its very nature beyond the sphere of material production proper. Just as the savage [*sic*] must wrestle with nature to satisfy his needs, to maintain and reproduce his life, so must civilized man, and he must do so in all forms of society and under all possible modes of production. This realm of natural necessity expands with his development, because his needs do too; but the productive forces to satisfy these expand at the same time. Freedom, in this sphere, can consist only in this, that socialized man, the associated producers, govern the human metabolism with nature in a rational way, bringing it under their collective control instead of being dominated by it as a blind power; accomplishing it with the least expenditure of energy and in conditions most worthy and appropriate for their human nature. But this always remains a realm of necessity. The true realm of freedom, the development of human powers as an end in itself, begins beyond it, though it can only flourish with this realm of necessity as its basis. The reduction of the working day is the basic prerequisite.[32]

Marx appeals here to the same basic idea of freedom which guides him elsewhere and which I have been defending in this dissertation (which is why I quoted part of this

---

[31] Arendt, *The Human Condition*.
[32] Marx, *Capital*, 3, 958–959.

passage in the Introduction). But some of the wording of this particular passage makes it difficult to reconcile with the possibility of free productive activity. Like Aristotle, Marx seems to be suggesting here that the problem with labor insofar as it aims at the maintenance and reproduction of life is that it is essentially done under the guise of necessity. A society of free people, accordingly, must be one in which this burden, though unavoidable in principle, is reduced temporally to a minimum, while the remaining time, the time beyond "the working day," is maximized.

This pessimistic assessment of labor as such is echoed, too, in contemporary liberal work on free time, as we already saw in Chapter III. Goodin *et al.* defend the importance of free time for autonomy by arguing specifically that "socially necessary" labor represents the negation of autonomy. This category of labor is narrower than Marx's target above, but the putative basis of its unfreedom is the same, and so is the implication. Inasmuch as labor is a form of subjection to necessity, we are supposed to conclude that freedom can begin only where labor ends, in the realm of "free time," at the end of the working day.[33]

We may also place Bertrand Russell's "In Praise of Idleness" loosely in the same school of thought.[34] There are some differences. Russell defines work not in terms of its necessity but (possibly tongue-in-cheek) as "altering the position of matter at or near the earth's surface relatively to other such matter" or "telling other people to do so."[35]

---

[33] Goodin et al., *Discretionary Time*.
[34] Bertrand Russell, "In Praise of Idleness," *Harper's Magazine* 165 (1932). See also John Maynard Keynes, "Economic Possibilities for Our Grandchildren," in *Essays in Persuasion* (New York: Harcourt Brace, 1932).
[35] Russell, "In Praise of Idleness," 553.

Unlike the liberals, moreover, though not unlike Marx, he identifies the source of avoidable toil in capitalist societies as what he calls "the divorce between the individual and the social purpose of production": the individual produces "for profit" (Russell here elides the distinction between worker and firm) but the social purpose of production is "the consumption of what he produces."[36] Like all of the philosophers mentioned above (in these texts or passages), though, he does not challenge the notion that labor in general, or at least most labor, is toil. The practical problem, once again, is supposed to be not the intrinsic unchoiceworthiness of work but—*taking for granted* its unchoiceworthiness—the length of the working day.

5.2

The expansion of free time outside work is one strategy for dealing with the problem of toil, but it is not the way I have been advocating so far, and in fact it fails to challenge the most important causes of unfree work in our society.

We can begin to see the limitations of the approach by reflecting more concretely on what sorts of activities would have to count as unfree according to the letter of these views. For instance, if the problem with unfree labor really is its necessity for life (perhaps at some minimal standard of living), then it might be possible autonomously to play basketball with one's friends but not to cook and eat dinner with one's friends, since cooking and eating are paradigm cases of necessary activity. Indeed, if—as seems plausible—some amount of physical exercise and socializing are also necessary for human life, then perhaps even the basketball game must be unfree. These activities

---

[36] Ibid., 558.

could only be free to the extent that they are superfluous (for life). The realm of freedom only begins with dessert. But all of this is absurd.

The problem, as I already argued in Chapter III, is that necessity alone is not enough for unfreedom; only *mere* necessity is. Insofar as one cooks and eats and plays basketball with friends not as a *mere* means to nourishment and health but also for its own sake, one does these things freely. The same goes for the production of durable material objects. Labor, then, is not unfree toil just by virtue of being labor. It becomes toil only in certain conditions. Further investigation *could* reveal that these conditions are more or less permanent features of human life. But this does not follow from the very idea of what labor is.

What happens now to the ideal of "free time?" The concept of free time implies *both* time away from work *and* time which is uniquely free. But for that reason, it ceases to have application if and when work comes to be done for its own sake and freely. It has application only in those benighted conditions in which the realm of work really does happen to be opposed to freedom. Theories which call predominantly for the expansion of free time hypostatize this condition. By opposing free time to labor *as such*, they treat the contingent and contestable unfreedom of toil as an uncontestable, ahistorical metaphysical fact.

Politically, this amounts to a kind of defeatism. As I have argued in the preceding sections, at least two important categories of contemporary toil—bullshit jobs and routine labor—stem from a historically specific and in principle alterable form of social organization. In light of such a diagnosis, to call only for the temporal reduction of the working day and the expansion of free time is to leave unchallenged the social

conditions which are the true cause of the unfreedom. At least in the first instance, our goal should not be to promote one pole of the opposition between work and free time at the expense of the other, but rather to transform the conditions which give rise to the opposition itself.[37]

I hasten to add that I am not suggesting that there is anything *wrong* with wanting or having more time outside work, so as to be able to do more things purely for their own sake or even for no reason at all. My point is that as far as the problem of toil is concerned, getting more "free time" in this sense alleviates the symptoms but does not cure the disease.

5.3

It might seem that my position diverges for the same reasons from so-called "anti-work" theory.[38] Theorists in this tradition describe the object of their critique as "work" or "labor" itself, and they speak of the importance of making time and social space for

---

[37] This is the direction which Marx takes in other writings, such as in the *Critique of the Gotha Program*, where he suggests that "[i]n a higher phase of communist society, […] the enslaving subordination of the individual to the division of labour, and therewith also the antithesis between mental and physical labour, has vanished; […] labour has become not only a means of life but life's prime want […]!" "Critique of the Gotha Programme," 531; compare *Economic and Philosophical Manuscripts of 1844*, 74. For my purposes, nothing compels us to go as far as identifying labor as life's *prime* want. That said, here as in his early work, I take Marx to be operating with such a broad, formal sense of 'labor' that the claim is not nearly as strong as it sounds. Allen Wood is quite good on these topics. See Allen W. Wood, *Karl Marx*, 2nd ed. (New York: Routledge, 2004), especially Chapters 2–4.

[38] See, e.g., André Gorz, *Paths to Paradise: On the Liberation from Work* (Boston: South End Press, 1985); Moishe Postone, *Time, Labor, and Social Domination: A Reinterpretation of Marx's Critical Theory* (Cambridge: Cambridge University Press, 1993); Kathi Weeks, *The Problem with Work: Feminism, Marxism, Antiwork Politics, and Postwork Imaginaries* (Durham, NC: Duke University Press, 2011); Frayne, *The Refusal of Work*.

activity beyond work. Invariably, however, they define 'work' and 'labor' in these contexts in a historically specific way, typically as some version of *paid labor.*[39] In opposing the realm of freedom to labor in some such historically specific sense, these theorists do not commit the mistake which I attributed to the authors above, of hypostatizing the contingent and historically specific unfreedom of toil. That said, it seems to me that using the *generic* words 'work' and 'labor' to refer to particular *species* of work and labor only invites confusion.

If we insist on using the generic term for a particular species of work, we leave ourselves with no clear and simple way of referring either to the genus or to the other species. Right out of the gates, this unnecessarily limits the apparent scope of our social theory, making it a poor conceptual tool for criticisms of work which do not rest entirely on its being paid labor. Most obviously, unpaid household labor is left out as an object of possible criticism, except insofar as it relates to paid labor. Anti-work theorists sometimes suggest that household labor is indeed "organized around" paid labor even if it is not itself paid.[40] Even if this is true, however, it is a very precarious thread on which to hang the much more straightforward truth that household labor is a form of work.[41]

---

[39] E.g., Weeks, *The Problem with Work*, 14, Frayne, *The Refusal of Work*, 19–20. In this usage, they follow Marx, who calls at various points not merely for a reduction of the working day but for the complete "abolition" of labor—where what he has in mind is labor in the historically specific form that it takes in capitalist society. See, e.g., *The German Ideology*, vol. 1, 220–221.

[40] E.g., Weeks, *The Problem with Work*, 14.

[41] Some Frankfurt School theorists, recognizing this problem, define labor more broadly in terms of the concept of "social contribution," where this is explicitly opposed to merely "private" instrumental activity. See, e.g., Friedrich Kambartel, "Arbeit und Praxis: Zu den begrifflichen und methodischen Grundlagen einer aktuellen politischen

Most importantly, though, anti-work theorists become unable, by virtue of their terminology, to distinguish in a clear and unmistakeable manner between the mere *reduction* of the working day and the *transformation* of the structure and content of work itself. To say that one is "anti-work" *simpliciter* and to propose to "abolish" work *simpliciter* almost irresistibly insinuates the strategy of reduction of the working day, the only apparent difference being that it is taken to an implausible extreme—implausible because, no matter how awesome our technological power may become, we will always remain materially dependent, socially interdependent beings with finite powers over nature. But plausible or implausible, this strategy leaves unchallenged the fundamental social problem, which is the social organization and content of work, not its relative extent. Now, given that they propose to abolish paid labor, anti-work theorists clearly *do* want to restructure and transform work. But their terminology stands in the way of clear comprehension of this goal, making their position seem somehow both less politically radical *and* less plausible than it really is.

Debatte" [Labor and Practice: On the Conceptual and Methodological Foundations of a Current Political Debate], *Deutsche Zeitschrift für Philosophie* 41, no. 2 (1993); and Rahel Jaeggi, "Pathologies of Work," *Women's Studies Quarterly* 45, no. 3 & 4 (2017). This definition is broader than the anti-work theorists' and can bring some household labor into the picture, but if the private–social distinction means anything at all, it makes the definition too narrow. The labor of personal care, including the household labor of people living alone, is also work, and there is no reason to think it lacks social significance. Consider, first, that such labor is also necessary for the society to maintain and reproduce itself, and second, that different people might have to do different amounts and types of personal care by virtue of differences of age, health, ability, and social position.

6. "NATURAL" UNFREEDOM?

I have chosen to focus on bullshit jobs and routine labor because the proliferation of these kinds of toil seems particularly clearly gratuitous and avoidable. Even if some toil were "naturally" necessary and unavoidable, however, its unfreedom would have practical implications.

6.1

Might some toil have "natural" rather than social causes?

Consider household chores such as cleaning, washing, and ironing. These chores are not pointless like bullshit jobs. Nor, when they are not done for pay, do these chores involve the kind of radical detachment from their purpose which characterizes routine labor. However, similarly to routine labor (and unlike other kinds of household and reproductive labor such as cooking and childrearing), they are characterized by a very limited scope for the exercise of skill: certainly the skills involved in this labor do not seem capable of indefinite development, affording ever new challenges and possibilities for meeting them in new ways. Moreover, these chores are profoundly repetitive. As Simone de Beauvoir writes:

> Few tasks are more similar to the torment of Sisyphus than those of the housewife; day after day, one must wash dishes, dust furniture, mend clothes that will be dirty, dusty, and torn again. The housewife wears herself out running on the spot; she does nothing; she only perpetuates the present; she never gains the sense that she is conquering a positive Good, but struggles indefinitely against Evil. It is a struggle that begins again every day.[42]

---

[42] Beauvoir, *The Second Sex*, 474. Compare Andrea Veltman, "The Sisyphean Torture of Housework," *Hypatia* 19, no. 3 (2004); and "Simone de Beauvoir and Hannah Arendt on Labor," *Hypatia* 25, no. 1 (2010).

This still does not sound like the description of an intrinsically choiceworthy way to spend one's time. Yet chores of this sort do not seem to derive their necessity from a particular social organization of labor; they would be necessary in any feasible form of society. Their necessity is not "social" but "natural."

Now, it should be emphasized that even if the necessity of household chores and similar labor is not a product of social conditions, it is not obvious that its toilsome character is similarly "natural." Certainly what makes it objectionable for Beauvoir is in the first instance not the content of the work *per se* but rather the fact that it when it takes up more or less one's entire existence, one is denied the possibility of participating in the further, less purely repetitive activities which one's labor makes possible.[43] In other words, once again, the most glaring problem is the division of labor which structures the work—not a given natural fact but a social condition.

Do repetitive, comparatively unskilled chores still have the character of toil if they occupy only a subordinate role in the life of the worker? Or can they be intrinsically choiceworthy in this context? This question—which hinges on substantive ethical questions about what has intrinsic value and what does not—is beyond the scope of this dissertation. For the sake of argument, though, I want to leave open the *possibility* that some such labor might be unavoidably toilsome, not by virtue of the form of its social organization but in its very nature. What would follow, given my account of freedom?

To begin with, even if we cannot *eliminate* certain kinds of toil altogether, it may

---

[43] Compare Simone de Beauvoir, *The Ethics of Ambiguity*, trans. Bernard Frechtman (New York: Philosophical Library, 1949), Chapter III, Section 2.

be the case that we can meaningfully *reduce* its scope in our lives, for instance, by technological means. According to one study, "the proportion of working-class women spending more than 9 hours a week in washing and ironing dropped from 61 per cent before the introduction of the washing machine to 24 per cent afterwards."[44] To the extent that this labor is a form of toil and therefore unfree, then, technology can be a means of relative liberation (though its emancipatory potential may not be realized if one form of toil is simply replaced by another).

The necessity of toil, even if it has "natural" rather than "social" causes, represents a threat to the ideal of a society of free people. I take it that the commitment to such a society is one which we can demand of one another and for which we can hold one another to account. We accordingly have a collective responsibility, in the face of ineliminable toil, to do what we can to reduce its scope in our lives to a negligible level; and we can make second-personal claims on one another to take steps toward the discharging of this collective responsibility. In other words, the "free time" approach, which I rejected in its general form above, is appropriate and indeed morally imperative in relation to the specific kind of toil whose necessity or unchoiceworthiness is not an avoidable product of contingent social conditions.

6.2

But what if some toil is not even meaningfully reducible given the limits of our

---

[44] Robert S. Lynd and Helen Merrell Lynd, *Middletown in Transition: A Study in Cultural Conflict* (New York: Harcourt Brace, 1937), quoted in Jonathan Gershuny, *Changing Times: Work and Leisure in Post-Industrial Society* (Oxford: Oxford University Press, 2000), 67, quoted in Goodin et al., *Discretionary Time*, 75n29.

technology?

In general, any social ideal must be beholden to some standard of feasibility. This is because a social ideal is more than a mere *wish*, a mere object of fancy. It must be capable of being practical, in the sense of serving as a possible *end* of action. But it can be practical in this sense only if something would count as actually doing something for the sake of it.[45] To the extent that our powers to eliminate or even reduce toil give out, however, it would seem that the ideal of freedom from such toil can no longer serve as a genuine end for us, for nothing would count as trying to bring about what one acknowledges to be impossible; nor would it make sense to demand of others that they try to do so. The ideal becomes a mere fantasy devoid of social or political significance.

But this argument is too quick. Even ineliminable, irreducible toil gives rise to two kinds of collective duties and corresponding second-personal claims. First, it is a mistake to treat the contingent limits on our powers simply as given. In general, these limits not only change over time but are themselves possible objects of *intentional* change. As Anton Ford writes:

> Practical limits are practical, not only in the sense that they are limits to what we can do, but also in the sense that they are limits that we can do something about. Things that are now impossible for some of us to do (e.g. make an international phone call), or for any of us to do (e.g. fly to Mars), may yet become possible through intentional human action. Nothing guarantees that they will become possible. They may well not. But the relevant limits are such that, in principle, we could change what they are.[46]

Human history is replete with expansions of our powers which have reduced our subjection, or at least our potential subjection, to constraints and compulsions which,

---

[45] Compare Kant, *The Metaphysics of Morals*, 6:213; Aristotle, *Nicomachean Ethics*, 1111a25.
[46] Anton Ford, "The Province of Human Agency," *Noûs* 52, no. 3 (2018): 716.

while their effects might have been distributed more equally, were not altogether eliminable before: witness the washing machine, and a whole variety of other tools more primitive and more sophisticated.[47] Like other tools, however, the washing machine did not simply fall from the heavens like mana. Its invention and adaptation for widespread use was the result of intentional action by human beings.

It is possible to pursue such expansions of our powers precisely for the sake of expanding human freedom. So the ideal of freedom can play a practical role even in relation to toil which is presently ineliminable and irreducible. What holds of first-personal practical reasoning, moreover, holds equally of second-personal claims and responsibilities. Even if no one is currently in a position to eliminate or reduce a certain kind of toil, we may still be in a position to make claims on one another to take steps toward putting its elimination or reduction within reach. Pablo Gilabert calls the responsibilities corresponding to claims of this nature "dynamic duties."[48]

The dynamic aspect of social ideals is just an application of a general point about action and practical reasoning. Action and deliberation for the sake of ends to which we currently lack sufficient means is not exceptional. I want to hitchhike from one city to another, but I am certain that no car driving to my destination will pass anywhere near me, so I hitch a ride to a third city, where I may—or may not (yet)—find a ride to my

---

[47] "Human freedom is younger than the species. Its most important features are only several thousand years old—an eyeblink in evolutionary history—but in that short time it has transformed the planet in ways that are as salient as such great biological transitions as the creation of an oxygen-rich atmosphere and the creation of multicellular life." Daniel C. Dennett, *Freedom Evolves* (New York: Penguin, 2003), 305.
[48] Pablo Gilabert, "Justice and Feasibility: A Dynamic Approach," in *Political Utopias: Contemporary Debates*, ed. Kevin Vallier and Michael Weber (Oxford: Oxford University Press, 2017).

destination. Or again, I want to bake a cake for a friend, but I lack the necessary skills, so I take a baking class, or watch a few videos on YouTube. At the limit, we may be so bereft of means that all we can do is to use our skills of perception and imagination *to look for* or *think of* means. But to do this, too, is already to take some means; it is already to take a first step in the direction of realizing our end. That end, therefore, has already become genuinely practical.

6.3

The second set of duties and claims to which ineliminable, irreducible toil can give rise concerns the just distribution of such toil.

Some toil, we are supposing, is an essential condition of human life as we know it; it is a necessary burden of social cooperation and essentially contributes to the benefits which people can expect from such cooperation (for instance, by sustaining and reproducing the members of the society over time). It therefore appears to be a fitting object of distributive justice. Other things equal, social arrangements in which some spend their lives performing toil on which their society depends so that others may lead lives free of this burden would seem to be to that extent unfair. Assuming the social necessity of some toil, each person has a duty to contribute their fair share of this burden. And this is a duty to which we can hold one another to account. Even utterly unavoidable toil, then, presents us with a practical problem: the problem of its just distribution.

I will not defend a concrete principle for the just distribution of toil here, but I do want to argue that toil should feature as a *distinct* object of distributive justice within

such a principle (or set of principles)—distinct, in particular, from money.[49] It might seem that justice in the distribution of money would be sufficient to ensure justice in the distribution of toil, given certain empirical assumptions. For if people who have a stronger preference for meaningful work or leisure can simply pay to be freed of the burden of toil by accepting a correspondingly lower wage, then the distribution of toil will simply reflect people's preferences.[50] Now, this argument is problematic for several reasons. First, in conditions of material inequality, the distribution of toil through the labor market reflects not just people's differential preferences, in the sense of their values, but also their differential ability to pay. When there is severe inequality, the distribution of toil can be expected *primarily* to reflect people's differential ability to buy their way out of it. Second, and relatedly, in reality the lowest-paid work tends also to be extremely burdensome toil, so the "trade-off" between money and meaningful work which this argument supposes is clearly a fantasy.[51]

The point I want to make, though, has to do with the special status that freedom should occupy in our social ideal. In a just society, certain goods, such as rights of political participation, cannot be bought and sold because only their independent, equal distribution can realize their distinct value, the value they have as expressions and

---

[49] For a book-length defense of the independent importance of the distribution of meaningful work on the one hand and toil on the other, see Paul Gomberg, *How to Make Opportunity Equal: Race and Contributive Justice* (Malden, MA: Blackwell, 2007). See also Andrea Veltman, *Meaningful Work* (Oxford: Oxford University Press, 2016).
[50] Nozick makes this sort of argument. Nozick, *Anarchy, State, and Utopia*, 246ff.
[51] Relatedly, see Rose's arguments that money and leisure are not perfectly substitutable. Julie L. Rose, "Money Does Not Guarantee Time: Discretionary Time as a Distinct Object of Distributive Justice," *The Journal of Political Philosophy* 22, no. 4 (2014); *Free Time* (Princeton: Princeton University Press, 2016), Chapter 4.

conditions of our free and equal citizenship. Not to have the right to vote, for instance, is to have a fundamentally lower social standing than one's fellow citizens, which is incompatible with the social equality that is a fundamental aspect of just social relations. Socioeconomic gains cannot in principle compensate for this loss of standing or the damage it does to social relations.

In a society of free and equal citizens, the freedom of the individual has a similar priority to other values. This is generally agreed when it comes to "basic liberties" such as freedom of conscience, freedom of speech, etc. These, too, cannot be justly forfeited in exchange for socioeconomic gains. But in light of the conception of freedom which I have defended, taking seriously the priority of freedom requires us also to take seriously the essential unfreedom of toil. A society some of whose citizens toil, in effect, for the sake of the others is one which does not take seriously some of its citizens' claims to freedom. This fundamentally fails to accord equal respect to their agency and, again, cannot be compensated by socioeconomic gains.

For this reason, the market, at least in an unrestricted form, would seem to be a singularly unsuitable mechanism for the just distribution of toil (quite aside from my arguments in Chapter IV). It is unsuitable for roughly the same reason it is an unsuitable mechanism for the distribution of jury duty: the type of labor being allocated has a special social significance which makes it non-commensurable both with other kinds of labor and with other goods such as money.

Whether or not the market can be a just way of distributing toil, though, it should now be clear at least that it can be quite misleading to speak of the unfreedom of toil as "natural," even in cases where it is unavoidable. In a society committed to the freedom of each of its members, "natural" unfreedom is social unfreedom.

CONCLUSION

In the course of defending the idea of freedom as activity for its own sake and spelling out some of its implications, several routes have led me to criticisms of the market. In Sections 7–9 of Chapter IV, I argued that pure exchange is morally objectionable on deontic grounds. In Section 3 of Chapter V, I argued, more tentatively, that the market economy is problematic on consequentialist grounds, because it tends to cause unfreedom by generating an artificial need for toil. Finally, in Section 6 of the same chapter, I argued that the labor market is an unjust mechanism for the distribution of toil in particular, because it is incompatible with the specificity of toil as a social burden given the priority of freedom.

I have said little to nothing, though, about the determinate practical implications of these criticisms. I have not considered with any specificity what an alternative to the market might look like. I have not considered the practical problem of how to get there from here. And, in particular, I have not considered the challenge of the feasibility of such a transition. The reality is that thinking through these issues in a serious way would take difficult empirical work (and not just economics) which is beyond the scope of this dissertation. My aim has been to lay some of the ethical foundations for this next step.

I recognize that this way of proceeding may seem objectionably utopian, in light of the essential practicality of social ideals. It may seem that until we have reason to believe that a putative social ideal is possible, nothing could count as trying to make

progress toward it, and so this ideal cannot really be practical; it can only be a wish, not an end. So it is a mistake to defend an abstract social ideal without first, or simultaneously, considering what it would look like concretely to realize this ideal and whether it is feasible.

While I would certainly have liked to be able to do more in this dissertation, I reject the charge of objectionable utopianism. First, analogously to the duties of reduction and redistribution in the case of potentially unavoidable toil, my arguments concerning the market might have "second-best" practical implications that fall short of abolishing it altogether. There might be policies possible within a market society which would significantly attenuate the unfreedom to which the market gives rise. Even if nothing strictly counted as a means to the abolition of the market, then, my deontic and consequentialist criticisms of the market could have practical significance.

A universal income not conditional on work, for instance, might help ensure that people working for money would at least no longer be working for the sake of access to basic necessities. This would directly reduce the degree of their unfreedom. It might also help indirectly reduce the toilsomeness of work. One wonders what it would take for a company like Amazon to employ workers in its warehouses if its would-be workers had an unconditional income sufficient to live well, and were thus in a condition more closely approximating genuine freedom of occupation.

I also want to resist the charge of objectionable utopianism at a more fundamental level. I accept the premise that a social ideal must be more than a wish; it must be practical, and this requires there to be something, here and now, which counts as trying to realize it. But as I argued at the end of Chapter V, the partly reflective work of trying to figure out how to achieve an end can itself be a means to that end. People

who are stuck in a cave may want to get out. They may have no idea how to do so, and no indication that it is possible, yet they may try to do so all the same. In part, their trying will take the form of thinking: for instance, thinking about where they are most likely to find an exit, or about how they should conduct their search, or about whether they are likely to be able to force their way out somehow. In part, it will consist in experimentally looking for a way out. Neither of these tasks presupposes the knowledge that they can in fact get out or the knowledge of how to get out. But the process does presuppose a clear-eyed apprehension of their predicament and an unrepressed desire to escape.

Milton and Rose Friedman frame what they take to be the fundamental problem of political theory as follows:

> The basic problem of social organization is how to co-ordinate the economic activities of large numbers of people. […] The challenge to the believer in liberty is to reconcile […] widespread interdependence with individual freedom. […] Fundamentally, there are only two ways of co-ordinating the economic activities of millions. One is central direction involving the use of coercion—the technique of the army and of the modern totalitarian state. The other is voluntary co-operation of individuals—the technique of the market place.[1]

This is an immensely influential picture.[2] I have argued that it is wrongheaded on two counts: the market is not a form of voluntary cooperation, and in principle there are possible forms of social coordination other than coercion and pure exchange. It may be that the Friedmans and their ilk simply consider the alternatives unrealistic on a large scale, and implicitly rule them out of consideration for that reason. Even so, however,

---

[1] Friedman and Friedman, *Capitalism and Freedom*, 12–13.
[2] For one of many other examples of this kind, see Rawls, *A Theory of Justice*, §42.

the claim that the market is consistent with individual freedom would not follow.

The truth is that this claim gives us a false sense of resolution, a delusion of reflective equilibrium. Once we see that it is mistaken, we recognize that capitalism does not satisfactorily solve "the basic problem of social organization." Whatever other wonderful things it may be, it is not a way of reconciling social interdependence and individual freedom. "The challenge to the believer in liberty," then, turns out to be more serious than liberal political theory would have us believe. But it is no genuine solution to adjust our ideal of freedom until it fits within the bounds of what currently seems feasible. Being stuck in a cave is no excuse for mistaking shadows on the wall for reality, or the fire for the sun.

BIBLIOGRAPHY

Adams, William. "Aesthetics: Liberating the Senses." In *The Cambridge Companion to Marx*, edited by Terrell Carver, 246–274. Cambridge: Cambridge University Press, 1991.

Anderson, Elizabeth. "What is the Point of Equality?" *Ethics* 109, no. 2 (1999): 287–337.

———. "Optional Freedoms." In *What's Wrong with a Free Lunch?*, edited by Joshua Cohen and Joel Rogers, 70–74. Boston, MA: Beacon Press, 2001.

———. "The Fundamental Disagreement Between Luck Egalitarians and Relational Egalitarians." *Canadian Journal of Philosophy* Supp. Vol. 36 (2010): 1–23.

———. *Private Government: How Employers Rule Our Lives (and Why We Don't Talk About It).* Princeton: Princeton University Press, 2017.

Anscombe, G. E. M. *Intention.* Cambridge, MA: Harvard University Press, 1963.

———. "Good and Bad Human Action." In *Human Life, Action and Ethics: Essays by G. E. M. Anscombe*, edited by Mary Geach and Luke Gormally, 195–206. Exeter: Imprint Academic, 2005.

Arendt, Hannah. *The Human Condition.* 2nd ed. Chicago: University of Chicago Press, 1998. 1958.

Aristotle. *Metaphysics.* Translated by W. D. Ross. In *The Complete Works of Aristotle*, Vol. 2, edited by Jonathan Barnes, 1552–1728. Princeton: Princeton University Press, 1984.

———. *Nicomachean Ethics.* Translated by W. D. Ross and J. O. Urmson. In *The Complete Works of Aristotle*, Vol. 2, edited by Jonathan Barnes, 1729–1867. Princeton: Princeton University Press, 1984.

———. *Politics.* Translated by B. Jowett. In *The Complete Works of Aristotle*, Vol. 2, edited by Jonathan Barnes, 1986–2129. Princeton: Princeton University Press, 1984.

Austin, J. L. *Sense and Sensibilia.* Oxford: Clarendon Press, 1962.

———. "A Plea for Excuses." In *Philosophical Papers*, edited by J. O. Urmson and G. J. Warnock, 175–204. Oxford: Oxford University press, 1979.

Babbage, Charles. *On the Economy of Machinery and Manufactures.* London, 1832. New York, 1963.

Barnes, Elizabeth. *The Minority Body: A Theory of Disability.* Oxford: Oxford University Press, 2016.

Bazargan, Saba. "Moral Coercion." *Philosophers' Imprint* 14, no. 11 (2014).

Beauvoir, Simone de. *The Ethics of Ambiguity.* Translated by Bernard Frechtman. New York: Philosophical Library, 1949. 1947.

———. *The Second Sex.* Translated by Constance Borde and Sheila Malovany-Chevallier. New York: Vintage, 2011. 1949.

Benn, S. I., and W. L. Weinstein. "Being Free to Act and Being a Free Man." *Mind* 80, no. 2 (1971): 194–211.

Berlin, Isaiah. "Two Concepts of Liberty." In *Four Essays on Liberty*, 118–172. Oxford: Oxford University Press, 1969.

Biggs, Joanna. *All Day Long: A Portrait of Britain at Work.* London: Profile Books, 2015.

Bovens, Luc. "Sour Grapes and Character Planning." *The Journal of Philosophy* 89 (1992): 57–78.

Bratman, Michael E. *Faces of Intention: Selected Essays on Intention and Agency.* Cambridge: Cambridge University Press, 1999.

Braverman, Harry. *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century.* 25th Anniversary ed. New York: Monthly Review Press, 1998.

Brennan, Jason, and Peter Martin Jaworski. "Markets Without Symbolic Limits." *Ethics* 125 (2015): 1053–1077.

Brudney, Daniel. "Community and Completion." In *Reclaiming the History of Ethics: Essays for John Rawls*, edited by Andrews Reath, Barbara Herman, Christine M. Korsgaard and John Rawls, 388–415. Cambridge: Cambridge University Press, 1997.

Business Roundtable. "Statement on the Purpose of a Corporation." 2019.

Buss, Sarah. "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints." *Ethics* 115 (2005).

Carter, Ian. *A Measure of Freedom.* Oxford: Oxford University Press, 1999.

Cohen, G. A. "The Structure of Proletarian Unfreedom." *Philosophy and Public Affairs* 12, no. 1 (1983): 3–33.

———. *Karl Marx's Theory of History: A Defense.* Expanded ed. Princeton: Princeton University Press, 2000.

Cranston, Maurice. *Freedom, A New Analysis.* London: Longman, Green & Co., 1953.

Csikszentmihaly, Mihaly. *Flow: The Psychology of Optimal Experience.* New York: Harper and Row, 1990.

Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability.* Cambridge, MA: Harvard University Press, 2006.

Deci, Edward L. "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology* 18 (1971): 105–115.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivations." *Psychological Bulletin* 125, no. 6 (1999): 627–668.

Defoe, Daniel. *Robinson Crusoe.* Oxford: Oxford University Press, 2007. 1719.

Dennett, Daniel C. *Freedom Evolves.* New York: Penguin, 2003.

Dworkin, Gerald. "Acting Freely." *Noûs* 4, no. 4 (1970): 367–383.

———. *The Theory and Practice of Autonomy.* Cambridge: Cambridge University Press, 1988.

Elster, Jon. *Sour Grapes: Studies in the Subversion of Rationality.* Cambridge: Cambridge University Press, 1983.

———. "Self-Realization in Work and Politics: The Marxist Conception of the Good Life." *Social Philosophy and Policy* 3, no. 2 (1986): 97–126.

Epictetus. *The Discourses of Epictetus.* Translated by Elizabeth Carter and Robin Hard. London: Everyman, 1995.

Evans, Will. "Ruthless Quotas at Amazon Are Maiming Employees." *The Atlantic*, 2019.

Feinberg, Joel. *Social Philosophy.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

———. *The Moral Limits of the Criminal Law, Volume 3: Harm to Self.* New York: Oxford University Press, 1986.

Fischer, John Martin. "Compatibilism." In *Four Views on Free Will*, edited by Robert Kane, John Martin Fischer, Derk Pereboom and Manuel Vargas. Malden: Blackwell, 2007.

Ford, Anton. "Action and Generality." In *Essays on Anscombe's Intention*, edited by Anton Ford, Jennifer Hornsby and Frederick Stoutland, 76–104. Cambridge, MA: Harvard University Press, 2011.

———. "The Province of Human Agency." *Noûs* 52, no. 3 (2018): 697–720.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68, no. 1 (1971): 5–20.

———. "Coercion and Moral Responsibility." In *Essays on Freedom of Action*, edited by Ted Honderich, 63–86. London: Routledge & Kegan Paul, 1973.

Frayne, David. *The Refusal of Work: Rethinking Post-Work Theory and Practice.* London: Zed Books, 2015.

Friedman, Milton, and Rose D. Friedman. *Capitalism and Freedom.* Chicago: University of Chicago Press, 1962.

Friedman, Richard. "Liberty Conceived as the Opposite of Slavery." In *Skepticism, Individuality, and Freedom: The Reluctant Liberalism of Richard Flathman*, edited by Bonnie Honig and David R. Mapel, 155–179. Minneapolis: University of Minnesota Press, 2002.

Gershuny, Jonathan. *Changing Times: Work and Leisure in Post-Industrial Society.* Oxford: Oxford University Press, 2000.

Geuss, Raymond. "Freedom as an Ideal." *Proceedings of the Aristotelian Society* Supp. Vol. 69 (1995): 87–100.

———. *History and Illusion in Politics.* Cambridge: Cambridge University Press, 2001.

Gilabert, Pablo. "Justice and Feasibility: A Dynamic Approach." In *Political Utopias: Contemporary Debates*, edited by Kevin Vallier and Michael Weber, 95–126. Oxford: Oxford University Press, 2017.

Gomberg, Paul. *How to Make Opportunity Equal: Race and Contributive Justice.* Malden, MA: Blackwell, 2007.

Goodin, Robert E. "Exploiting a Situation and Exploiting a Person." In *Modern Theories of Exploitation*, edited by Andrew Reeve, 166–200. London: Sage, 1987.

Goodin, Robert E., James Mahmud Rice, Antti Parpo, and Lina Eriksson. *Discretionary Time: A New Measure of Freedom.* Cambridge: Cambridge University Press, 1998.

Gorz, André. *Paths to Paradise: On the Liberation from Work.* Boston: South End Press, 1985.

Graeber, David. *Bullshit Jobs: A Theory.* New York: Simon & Schuster, 2018.

Grant, Ruth W. *Strings Attached: Untangling the Ethics of Incentives.* Princeton: Princeton University Press, 2012.

Greenfield, Lauren. *Generation Wealth.* London: Phaidon, 2017.

Haksar, Vinit. "Coercive Proposals [Rawls and Gandhi]." *Political Theory* 4, no. 1 (1976): 65–79.

Hayek, Friedrich A. *The Constitution of Liberty.* Chicago: University of Chicago Press, 1960.

Herman, Barbara. "Leaving Deontology Behind." In *The Practice of Moral Judgment*, 208–240. Cambridge, MA: Harvard University Press, 1993.

Hobbes, Thomas. *Leviathan.* Indianapolis: Hackett, 1994.

———. "Of Liberty and Necessity." In *Hobbes and Bramhall on Liberty and Necessity*, edited by Vere Chappell, 15–42. Cambridge: Cambridge University Press, 1999.

Humboldt, Wilhelm von. *The Sphere and Duties of Government.* Translated by Joseph Coulthard. London: John Champan, 1854.

Hume, David. *An Enquiry Concerning Human Understanding.* Second ed. Indianapolis: Hackett, 2011.

Hurka, Thomas. "Why Value Autonomy?". *Social Theory and Practice* 13, no. 3 (1987): 361–382.

Jaeggi, Rahel. *Alienation.* Translated by Frederick Neuhouser and Alan E. Smith. New York: Columbia University Press, 2014.

———. "Pathologies of Work." *Women's Studies Quarterly* 45, no. 3 & 4 (2017): 59–76.

Johnson, Harriet McBryde. "Unspeakable Conversations." *New York Times Magazine*, February 16 2003.

Julius, A. J. "The Possibility of Exchange." *Politics, Philosophy & Economics* 12, no. 4 (2013): 361–374.

———. "A Form of Unfreedom." (Manuscript).

———. "Suppose We Had Produced as Humans." (Manuscript).

Kambartel, Friedrich. "Universalität als Lebensform." In *Philosophie der humanen Welt*, 14–26. Frankfurt: Suhrkamp, 1989.

———. "Arbeit und Praxis: Zu den begrifflichen und methodischen Grundlagen einer aktuellen politischen Debatte." *Deutsche Zeitschrift für Philosophie* 41, no. 2 (1993): 239–249.

Kant, Immanuel. *Critique of Pure Reason.* Translated by Norman Kemp Smith. New York: Palgrave Macmillan, 1929.

———. *Critique of Practical Reason.* Edited and Translated by Mary Gregor. In *Practical Philosophy*, 133–271. Cambridge: Cambridge University Press, 1996.

———. *The Metaphysics of Morals.* Edited and Translated by Mary Gregor. In *Practical Philosophy*, 353–603. Cambridge: Cambridge University Press, 1996.

———. *Lectures on Ethics.* Translated by Peter Heath. Cambridge: Cambridge University Press, 1997.

———. *Groundwork of the Metaphysics of Morals.* Translated by Mary Gregor. Cambridge: Cambridge University Press, 1998.

Keynes, John Maynard. "Economic Possibilities for Our Grandchildren." In *Essays in Persuasion*, 358–373. New York: Harcourt Brace, 1932.

Killmister, Suzy. "The Woody Allen Puzzle: How 'Authentic Alienation' Complicates Autonomy." *Noûs* 49, no. 4 (2015): 729–747.

———. *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance.* Routledge, 2018.

Knight, Frank. *Freedom and Reform.* New York: Harper & Brothers, 1947.

Korsgaard, Christine M. "Two Distinctions in Goodness." *The Philosophical Review* 92, no. 2 (1983): 169–195.

———. "Kant's Formula of Humanity." In *Creating the Kingdom of Ends*, 106–132. Cambridge: Cambridge University Press, 1996.

———. "The Right to Lie: Kant on Dealing with Evil." In *Creating the Kingdom of Ends*, 133–158. Cambridge: Cambridge University Press, 1996.

Kristjánsson, Kristján. *Social Freedom: The Responsibility View.* Cambridge: Cambridge University Press, 1996.

Laurence, Ben. "An Anscombean Approach to Collective Action." In *Essays on Anscombe's Intention*, edited by Anton Ford, Jennifer Hornsby and Frederick Stoutland. Cambridge, MA: Harvard University Press, 2010.

———. *Agents of Change: Political Philosophy as Practical Reasoning.* Manuscript.

Locke, John. *An Essay Concerning Human Understanding.* Oxford: Oxford University Press, 1975. 1689.

Lynd, Robert S., and Helen Merrell Lynd. *Middletown in Transition: A Study in Cultural Conflict.* New York: Harcourt Brace, 1937.

MacCallum, Gerald C., Jr. "Negative and Positive Freedom." *The Philosophical Review* 76, no. 3 (1967): 312–334.

Maguire, Barry, and Brookes Brown. "Markets, Interpersonal Practices, and Signal Distortion." *Philosophers' Imprint* 19, no. 14 (2019).

Marx, Karl. *Grundrisse: Foundations of the Critique of Political Economy (Rough Draft).* Translated by Martin Nicolaus. London: Penguin, 1973.

———. *Capital: A Critique of Political Economy.* Vol. 1. Translated by Ben Fowkes. New York: Vintage, 1977.

———. *Capital: A Critique of Political Economy.* Vol. 3. Translated by David Fernbach. London: Penguin, 1981.

———. *Economic and Philosophical Manuscripts of 1844.* Translated by Martin Milligan. New York: Prometheus Books, 1988.

———. "Critique of the Gotha Programme." In *Karl Marx: Selected Writings*, edited by David McLellan, 610–616. Oxford: Oxford University Press, 2000.

———. "On James Mill." In *Karl Marx: Selected Writings*, edited by David McLellan, 124–133. Oxford: Oxford University Press, 2000.

Marx, Karl, and Friedrich Engels. *The German Ideology.* 3rd ed. Moscow: Progress Publishers, 1976.

McGregor, Joan. "Bargaining Advantages and Coercion in the Market." *Philosophy Research Archives* 14 (1988–1989): 23–50.

Mele, Alfred R. *Autonomous Agents: From Self-Control to Autonomy.* New York: Oxford University Press, 1995.

Mill, John Stuart. *On Liberty.* In *On Liberty and The Subjection of Women*, edited by Alan Ryan, 1–129. London: Penguin, 2006.

———. *The Subjection of Women.* In *On Liberty and The Subjection of Women*, edited by Alan Ryan, 131–243. London: Penguin, 2006.

Miller, David. "Constraints on Freedom." *Ethics* 94 (1983): 66–86.

Nozick, Robert. "Coercion." In *Philosophy, Science, and Method*, edited by S. Morgenbesser, P. Suppes and M. White. New York, 1969.

———. *Anarchy, State, and Utopia.* New York: Basic Books, 1974.

———. *Philosophical Explanations.* Cambridge, MA: Harvard University Press, 1981.

Nussbaum, Martha. *Women and Human Development: The Capabilities Approach.* Cambridge: Cambridge University Press, 2000.

———. "Symposium on Amartya Sen's Philosophy: 5. Adaptive Preferences and Women's Options." *Economics and Philosophy* 17, no. 1 (2001): 67–88.

O'Neill, Onora. "Between Consenting Adults." *Philosophy and Public Affairs* 14, no. 3 (1985): 252–277.

Oppenheim, Felix E. *Dimensions of Freedom.* New York: St. Martin's Press, 1961.

Pallikkathayil, Japa. "Deriving Morality from Politics: Rethinking the Formula of Humanity." *Ethics* 121, no. 1 (2010): 116–147.

———. "The Possibility of Choice: Three Accounts of the Problem with Coercion." *Philosophers' Imprint* 11, no. 16 (2011).

Pettit, Philip. *Republicanism: A Theory of Freedom and Government.* Oxford: Oxford University Press, 1997.

Plamenatz, J. P. *Consent, Freedom and Political Obligation.* London: Oxford University Press, 1938.

Postone, Moishe. *Time, Labor, and Social Domination: A Reinterpretation of Marx's Critical Theory.* Cambridge: Cambridge University Press, 1993.

Radcliffe Richards, Janet. "Nephrarious Goings On: Kidney Sales and Moral Arguments." *The Journal of Medicine and Philosophy* 21 (1996): 375–416.

Rawls, John. *A Theory of Justice.* Cambridge, MA: Harvard University Press, 1971.

———. *Justice as Fairness: A Restatement.* Cambridge, MA: Harvard University Press, 2001.

Raz, Joseph. *The Morality of Freedom.* Oxford: Oxford University Press, 1986.

Ripstein, Arthur. "Rationality and Alienation." *Canadian Journal of Philosophy* Supp. Vol. 15 (1989): 449–466.

———. *Force and Freedom: Kant's Legal and Political Philosophy.* Cambridge, MA: Harvard University Press, 2009.

Rose, Julie L. "Money Does Not Guarantee Time: Discretionary Time as a Distinct Object of Distributive Justice." *The Journal of Political Philosophy* 22, no. 4 (2014): 438–457.

———. *Free Time.* Princeton: Princeton University Press, 2016.

Rousseau, Jean-Jacques. *Emile: or, On Education.* Translated by Allan Bloom. Basic Books, 1979.

Russell, Bertrand. "In Praise of Idleness." *Harper's Magazine* 165 (1932): 552–559.

Sachs, Benjamin. "Why Coercion is Wrong When It's Wrong." *Australasian Journal of Philosophy* 91, no. 1 (2013): 63–82.

Sandel, Michael J. "What Money Can't Buy: The Moral Limits of Markets." In *Tanner Lectures on Human Values*, edited by Grethe Peterson, 89–122. Salt Lake City: University of Utah Press, 2000.

———. *What Money Can't Buy: The Moral Limits of Markets.* New York: Farrar, Straus and Giroux, 2012.

Satz, Debra. *Why Some Things Should Not Be for Sale: The Moral Limits of Markets.* Oxford: Oxford University Press, 2010.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame.* Cambridge, MA: Harvard University Press, 2008.

Schmidt, Andreas T. "Abilities and the Sources of Unfreedom." *Ethics* 127 (2016): 179–207.

Schwartz, Adina. "Meaningful Work." *Ethics* 92, no. 4 (1982): 634–646.

Sen, Amartya. "Justice: Means vs. Freedoms." *Philosophy and Public Affairs* 19, no. 2 (1990): 111–121.

———. *Commodities and Capabilities.* New Delhi: Oxford University Press, 1999.

———. *Collective Choice and Social Welfare.* Expanded ed. Cambridge, MA: Harvard University Press, 2017.

Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations.* Chicago: University of Chicago Press, 1976.

Steiner, Hillel. "Individual Liberty." *Proceedings of the Aristotelian Society* 75 (1974–1975): 33–50.

Strawson, P. F. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*, 1–28. London: Routledge, 2008.

Swift, Adam. *Political Philosophy: A Beginners' Guide for Students and Politicians.* Cambridge: Polity Press, 2006.

Tannenbaum, Julie. "Categorizing Goods." *Oxford Studies in Metaethics* 5 (2010): 257–294.

Taylor, Charles. "What's Wrong with Negative Liberty." In *Philosophical Papers*, Vol. 2, 211–229. Cambridge: Cambridge University Press, 1985.

Taylor, Frederick W. *Principles of Scientific Management.* In *Scientific Management*. New York and London, 1947.

Terkel, Studs. *Working: People Talk About What They Do All Day and How They Feel About What They Do.* New York: Random House, 1972.

Tugendhat, Ernst. *Self-Consciousness and Self-Determination.* Translated by Paul Stern. Cambridge, MA: The MIT Press, 1986.

Van Parijs, Philippe. *Real Freedom for All: What (If Anything) Can Justify Capitalism?* Oxford: Oxford University Press, 1995.

Velleman, J. David. *The Possibility of Practical Reason.* Oxford: Oxford University Press, 2000.

Veltman, Andrea. "The Sisyphean Torture of Housework." *Hypatia* 19, no. 3 (2004): 121–143.

———. "Simone de Beauvoir and Hannah Arendt on Labor." *Hypatia* 25, no. 1 (2010): 55–78.

———. *Meaningful Work.* Oxford: Oxford University Press, 2016.

Vogler, Candace. *Reasonably Vicious.* Cambridge, MA: Harvard University Press, 2002.

Watson, Gary. "Free Agency." *The Journal of Philosophy* 72, no. 8 (1975): 205–220.

Weeks, Kathi. *The Problem with Work: Feminism, Marxism, Antiwork Politics, and Postwork Imaginaries.* Durham, NC: Duke University Press, 2011.

Wertheimer, Alan. *Coercion.* Princeton: Princeton University Press, 1987.

White, Stephen J. "On the Moral Objection to Coercion." *Philosophy and Public Affairs* 45, no. 3 (2017): 199–231.

Whiting, Jennifer. "Eudaimonia, External Results, and Choosing Virtuous Actions for Themselves." *Philosophy and Phenomenological Research* 65, no. 2 (2002): 270–290.

Williams, Bernard. "From Freedom to Liberty: The Construction of a Political Value." *Philosophy and Public Affairs* 30 (2001).

Wittgenstein, Ludwig. *Philosophical Investigations.* Translated by G. E. M. Anscombe. 3rd ed. Malden, MA: Blackwell, 2001.

Wolf, Susan. "Asymmetrical Freedom." *The Journal of Philosophy* 77, no. 3 (1980): 151–166.

———. "Sanity and the Metaphysics of Responsibility." In *Responsibility, Character and the Emotions*, edited by Ferdinand Schoeman. Cambridge: Cambridge University Press, 1988.

———. *Freedom within Reason.* New York: Oxford University Press, 1990.

Wood, Allen W. *Karl Marx.* 2nd ed. New York: Routledge, 2004. 1981.

Yaffe, Gideon. "Indoctrination, Coercion and Freedom of the Will." *Philosophy and Phenomenological Research* 67, no. 2 (2003): 335–356.

Zimmerman, David. "Coercive Wage Offers." *Philosophy and Public Affairs* 10, no. 2 (1981): 121–145.