Autonomy and Coercion

Pascal Brixel
Clemson University

Abstract: I defend an account of coercion which explains the unfreedom of coerced action and the wrongfulness of coercion. Unlike traditional accounts, I begin with what coercive threats and mere offers have in common and what distinguishes them both from ordinary rational persuasion. Both threats and offers are species of incentivization, premised on a distinctively extrinsic form of motivation. But activity motivated in this extrinsic way, I argue, is essentially less than fully autonomous, for a familiar Kantian reason: because it is not underwritten without qualification by the agent's own practical reason. Fundamentally, I conclude, coercion is wrong because it is a species of incentivization, and incentivization in general compromises autonomy. After defending this claim, I show that it is compatible with a reasonable account of the moral difference between threats and offers. I close by outlining the implications of my view for the moral standing of the market.

Some ways of getting other people to do what we want are compatible with their freedom or autonomy; others are not. Coercion is not, and that explains what is most fundamentally wrong with it. But how exactly does coercion compromise a person's freedom or autonomy? When I speak of "coercion" in this paper, I mean not physical coercion, which operates by direct manipulation of another person's body, but what is sometimes called "volitional coercion," which operates by means of threats. Now, if I threaten to have you beaten up unless you do what I want, the choice is still yours: comply and avoid punishment, or refuse and be punished. If you choose to comply, what makes your choice less than fully free or autonomous?

Much ink has been spilled on attempts to solve this problem by focusing in the first instance on what distinguishes threats of punishment from offers of reward. While threats of punishment wrongfully compromise a person's autonomy, it is commonly assumed that offers of reward are normally neither autonomy-compromising nor wrongful. I argue that this assumption is mistaken. In fact, I argue, we can only bring into view why coercion compromises autonomy or what is most fundamentally wrong with it by focusing in the first

instance on what threats and offers have in common, and what distinguishes them both from other ways of influencing people.

Both threats and offers are species of incentivization; they operate by giving someone an incentive, a type of motive that is, in a distinctively radical sense, extrinsic to what it motivates. But action motivated in this extrinsic way, I argue, is necessarily less than fully autonomous, for a familiar Kantian reason: because it is not fully underwritten by the agent's own practical reason. Incentivization in general thus amounts to an attempt to control another person's conduct in a way which is incompatible with their self-determination, and it normally constitutes a *pro tanto* moral wrong for this reason. I conclude that the wrong of coercion is fundamentally a species of this more generic wrong.

The real moral difference between threats and offers, then, is not that the former are autonomy-compromising and objectionable while the latter are autonomy-preserving and morally innocuous. Instead, I argue, what distinguishes threats from offers is that the former involve a distinctively unilateral and committal, rather than partial and opportunistic, form of control of another's conduct. This remains, however, a distinction within heteronomy.

I close the paper by outlining the implications of my view for the moral standing of the market. I argue that the view has mixed implications. On the one hand, my account of the moral difference between threats and offers vindicates a traditional defense of the market over coercive alternatives such as feudalism and slavery. On the other hand, my criticism of incentivization in general suggests that the market as we know it is *pro tanto* morally objectionable on the grounds that it is incompatible with full individual autonomy and with social relations premised on full respect for one another as independent rational agents.

Views similar to the one I defend in this paper, sometimes associated with the label "freedom as independence," have been put forward by A.J. Julius and Nicholas Vrousalis.[1] By and large, though, this sort of approach has not yet established itself in the literature on the same footing as the Kantian, moralistic, and psychologistic alternatives. My main contributions in this paper consist in showing that the approach can be defended on the basis of traditional Kantian premises, that it solves some deep and intractable problems concerning coercion which the mainstream theories fail to solve, and that it is

[1] See, e.g., A. J. Julius, "The Possibility of Exchange," *Politics, Philosophy & Economics* 12, no. 4 (2013); Nicholas Vrousalis, "The Capitalist Cage: Structural Domination and Collective Agency in the Market," *Journal of Applied Philosophy* 38, no. 1 (2021): 41.

compatible with a reasonable account of the moral difference between threats and offers.

I. WHAT DO WE WANT FROM A THEORY OF COERCION?

If possible, a theory of coercion should explain and vindicate three pieces of common sense about coercion: that coerced action is distinctively unfree, that coercion is normally wrong, and that coercion is normally wrong *because* it compromises freedom.

To act under coercion is to act unfreely. Indeed, coerced action is a paradigm case of unfree action. The mugging victim who is threatened with murder unless she hands over her wallet does not hand over her wallet freely. The slave who is threatened with a beating unless he cuts the crops does not cut the crops freely. These agents do not choose their actions voluntarily but are forced to do them. They are determined not by their own wills but by the will of another person. Even in special cases where the coerced action is in its victim's own interest or where it can be said to expand the victim's own freedom globally or in the long run, to suggest that the coerced action itself is done freely would be to make a mockery of the idea of freedom. The first desideratum for a theory of coercion, then, is that the theory explain the distinctive unfreedom of coerced action. The theory is incomplete if it fails to do this.

To coerce another person is to make them unfree. It is also to wrong them, at least normally—that is to say, in the absence of special defeating considerations—and *pro tanto*. A victim of coercion normally has grounds not just for lamentation about her misfortune. A second desideratum for a theory of coercion is that the theory explain this distinctive wrongfulness of coercion.

The unfreedom of coerced action and the wrongfulness of coercion are intimately connected. Coercing people seems wrong *because* it is makes them unfree. This explanatory relation distinguishes both the relevant kind of unfreedom and the relevant kind of wrong. On the one hand, the relevant kind of unfreedom is one which must be capable of having this distinctive normative significance. On the other hand, the relevant kind of wrong must be dependent on considerations of freedom and unfreedom. A third desideratum for a theory of coercion, then, is that the theory account for these connections. In particular, our theory ought to identify a kind of unfreedom whose violation we can reasonably understand as normally *pro tanto* wrong, and it ought to explain the wrongfulness of coercion on the basis of the way in which it makes its victim unfree.

Perhaps further inquiry will force us to give up on these desiderata to some extent. Pre-theoretically, however, it seems hard to deny that this is what we would ideally like from a theory of coercion: an explanation of why coerced action is unfree, which in turn grounds an explanation of why coercion is normally distinctively wrong.

II. KANTIANISM, MORALISM, AND PSYCHOLOGISM

The main theories of coercion in the literature fall into three categories: traditional Kantian theories grounded in the idea that coercion subverts another's practical reason, moralistic theories grounded in the idea that coercion involves an independent violation of another's rights, and psychologistic theories grounded in the idea that coercion involves a form of motivation which tends to give rise to a negative psychological reaction in the coerced agent. None of these approaches, I argue, successfully explains the wrongfulness of coercion by explaining the unfreedom of coerced action.

A. Traditional Kantianism

According to the old Kantian formula, what makes coercion objectionable is that it is a way of using another person as a mere means. Traditional Kantian accounts of this idea, however—such as defended by Christine Korsgaard, Onora O'Neill, and Barbara Herman—face an instructive difficulty, connected with the fact that when an action is coerced by means of threats rather than immediate physical force, it still constitutes a genuine exercise of agency on the part of the coerced agent.

Kantian self-determination is intimately bound up with our power of agency, or practical reason. As rational agents, we are not mere automata reacting mechanically to stimuli, but practical reasoners. A practical reasoner is capable of making up their own mind about what they have good reason to do. An action is self-determined or autonomous, accordingly, only if it is chosen ultimately on the basis of such a judgment on the part of the agent themselves: in other words, only if the agent does what they are doing because they take themselves to have good reason to do it.

In our dealings with other people, we are morally bound to respect their distinctive status as rational agents—as Kantians put it, to treat them as ends in themselves rather than as mere means.[2] This means, in particular, that

[2] In Kant's words, the 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,

when we try to get others to do what we want, we must do so in a way which is compatible with their doing it autonomously, on the basis of their own judgment that they have good reason to do it.

As Korsgaard suggests, ordinary rational persuasion is a paradigm case of this autonomy-preserving kind of interpersonal influence:

> 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.[3]

When I successfully persuade you to do something by drawing your attention to the existence of good reasons for doing that thing, you act in some sense under my influence. Nevertheless, your action proceeds ultimately from an unqualified exercise of your own practical reason and can thus be autonomous. By contrast, when I coerce you to do something, I do not give you a chance to choose whether to do it by making up your own mind about whether it is worth doing. Thus, O'Neill says that coercion "denies [its victims] the choice between consent and dissent";[4] Korsgaard characterizes it as a way of "taking a decision out of someone's hands," hence a way of treating her "as a mere means, a thing, a tool."[5]

On the Kantian picture, then, coercion is wrong for the following reason. As a practical reasoner, each of us has a right to autonomy: that is, a right to be the ultimate arbiters of what we ourselves are to do. But coercion is a way of influencing another person's will which usurps this authoritative role of their own practical reason in determining their activity. Coercion is wrong,

---

always at the same time as an end, never merely as a means." Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1998), 429, italics removed. As Kantians have traditionally interpreted it, "humanity" here means the aspect of our practical reason which consists in the ability to determine our own ends. For a defense of this reading, see Christine M. Korsgaard, "Kant's Formula of Humanity," in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), especially 110–14.

[3] Korsgaard, "The Right to Lie: Kant on Dealing with Evil," 142. Compare Barbara Herman, "Leaving Deontology Behind," in *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993).

[4] Onora O'Neill, "Between Consenting Adults," *Philosophy and Public Affairs* 14, no. 3 (1985): 262.

[5] Korsgaard, "The Right to Lie.", 138, 142.

then, because it compromises the other's autonomy—which is precisely the form of explanation that we want from a theory of coercion.

This account, moreover, appears to distinguish coercion not just from ordinary rational persuasion but also from another form of interpersonal influence commonly thought to be normally autonomy-preserving and morally innocuous: offers of reward. If I offer you money in exchange for some service, for example, I try to influence your will, but I do not take the decision out of your hands or deny you the opportunity to consent or dissent. On the contrary, all I do is present a proposal for your consideration. It is then up to you to choose to take it or leave it, on the basis of your own practical reasoning.

In its traditional forms, however, the Kantian approach faces serious difficulties. Things are clear enough with respect to physical coercion, which literally bypasses its victim's practical reason altogether. If someone kidnaps me by grabbing me and forcing me into a car, entering the car is in no meaningful sense my own action. It is trivially true that I do not enter the car because I take myself to have good reason to do so. Volitional coercion, however, which operates by means of threats, does not altogether bypass its victim's practical reason. Faced with a choice between handing my wallet over to a mugger or allowing myself to be shot, I can still reflect on the options open to me and make a decision based on what will best advance my own ends. If I comply and hand over my wallet, this is a genuine exercise of my practical reason. It is thus unclear in what sense my coercer genuinely takes the decision out of my hands, or deprives me of the power to consent or dissent. Call this the problem of the inalienability of agency.

Some tempting responses to this problem fail immediately. For example, O'Neill 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."[6] But a mugging victim who hands over her wallet is quite clearly pursuing ends: for example, the end of survival. Arthur Ripstein, in his more recent reconstruction of Kant's political philosophy, suggests that victims of coercion do pursue ends, but only ends set for them by their coercers:

> 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.[7]

---

[6] O'Neill, "Between Consenting Adults," 262.

[7] Arthur Ripstein, *Force and Freedom: Kant's Legal and Political Philosophy* (Cambridge, MA: Harvard University Press, 2009), 36.

But if a slave is ordered to do something on pain of being deprived of dinner, and she complies, her end of getting to eat dinner is not set for her by her master. David Zimmerman, who is sympathetic to the Kantian approach, suggests that victims of coercion are restricted to the pursuit of only a narrow, short-term range of goods such as immediate survival or the avoidance of pain.[8] But this need not be true. For one thing, it is possible to coerce someone by threatening a long-term good which they hold dear; for another, in the very act of pursuing a good such as survival, one can also be pursuing the many further ends to which survival is a means.

Sarah Buss, who has pressed this general criticism of the Kantian tradition in greater detail, summarizes the challenge thus:

> To put it somewhat crudely, 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.[9] And this would seem to depend on whether *she* determines her response to the considerations that figure in her reasoning […].

Since it seems that an agent determines her response to these considerations *whenever* she acts intentionally, the criterion of Kantian self-determination seems too weak to rule out volitional coercion. The problem, it appears, is the difficulty of avoiding an all-or-nothing conception of the authority of the agent's own practical reason over her activity. What is missing from the traditional Kantian picture is a clear understanding of coerced action as a genuine but privative exercise of agency: something less than fully self-determined practical reasoning, but something more than purely passive movement.

In the absence of such an understanding, the traditional Kantian explanation of the wrongfulness of coercion collapses as well. Coercion is supposed to be wrong because it is a way of failing to respect another person's status as a practical reasoner. But if coercion does not actually bypass or subvert the other's practical reason, we have no grounds for saying this. In its traditional forms, then, the Kantian approach begs all the important questions. It does not explain the sense in which coerced action is unfree, and it does not explain why coercion is wrong.

---

[8] David Zimmerman, "Coercive Wage Offers," *Philosophy and Public Affairs* 10, no. 2 (1981): 130.

[9] Sarah Buss, "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints," *Ethics* 115 (2005): 214.

B. Moralism

Many theorists of coercion, including a number of Kantians, have tried to make progress by defending an essentially moralized conception of coercion. According to this moralistic approach—defended by Alan Wertheimer, T.M. Scanlon, Japa Pallikkathayil, Arthur Ripstein, Stephen White, and many others—what is distinctively wrong with coercion is that the coercer does something independently wrong—that is, wrong on grounds independent of the fact that it is coercive—to his victim.

On the most straightforward and most popular version of moralism, the suggestion is that the coercive threat is wrong because the threatened action itself would be wrong on independent grounds. As Pallikkathayil says, the wrong of coercion is "parasitic on the wrongfulness of acting on the intention being announced."[10] For example, the mugger has no right to shoot his victim and would violate her rights if he did so. As Ripstein puts it, he "is offering something that he has no right to offer."[11] But for that reason, the mugger also wrongs his victim merely by threatening to shoot her. Once again, the contrast with offers seems instructive. There is nothing generally wrong with inducing another person to do something by offering to give them money in exchange, because there is nothing independently wrong with giving someone money. In this way, it seems we can explain what is distinctively wrong with typical threats of punishment compared with typical offers of reward.

The most obvious problem with this type of account of coercion is that it does nothing to explain the unfreedom of coerced action. It is hard to shake the naive 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

---

[10] Japa Pallikkathayil, "The Possibility of Choice: Three Accounts of the Problem with Coercion," *Philosophers' Imprint* 11, no. 16 (2011): 18. See also, e.g., Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), 262–64; Alan Wertheimer, *Coercion* (Princeton: Princeton University Press, 1987); T. M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, MA: Harvard University Press, 2008); Ripstein, *Force and Freedom*.; and Benjamin Sachs, "Why Coercion is Wrong When It's Wrong," *Australasian Journal of Philosophy* 91, no. 1 (2013). For more complicated moralistic theories, see, e.g., 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).

[11] Ripstein, *Force and Freedom*, 132.

seems to be neither here nor there. Granted that the mugger violates his victim's rights by threatening to shoot her, it still remains the case that she can now decide for herself how to respond to this threat. Why is this choice unfree? In the absence of an answer to this question, the theory of coercion remains importantly incomplete.[12]

Perhaps it is a mistake to see moralists as trying to offer a complete theory of coercion. Perhaps all they are after is an explanation of the wrongfulness of coercion.[13] Even measured by this more limited ambition, however, moralistic theories fall short, for the moral problem cannot be disentangled from the conceptual problem of the unfreedom of coerced action. Coercion is not just any old way of wronging someone. It is a way of wronging someone by making them unfree. But as we have seen, according to the moralistic approach, coercion is fundamentally wrongful not because it makes someone unfree but because it involves some independent wrong, such as threatening to violate someone's rights. So the account necessarily explains the wrongfulness of coercion in the wrong way.

In the literature, the problem tends to be understood primarily as a threat of extensional inadequacy. It is well known that moralists tend to have difficulties explaining the wrongfulness of coercive acts in those special cases where the threatened action is not independently wrong. For example, if I threaten to reveal embarrassing information about you unless you do what I want, then I wrong you even if I have a right to reveal the information and would not be wronging you by revealing it *per se*. Likewise, if an employer threatens to fire his employee unless she has sex with him, he wrongs her even if he in fact has a right to fire her and would not be wronging her simply by

---

[12] Stephen White tries to address this problem by suggesting that the coercer cannot "legitimately" and "in good faith" see himself as doing anything but interfering with his victim's "ability to respond to [her] situation on [her] own terms." White, "On the Moral Objection to Coercion," 230. But at the end of the day, White concedes 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." *Ibid.*, 231. And this concession seems fatal. For surely it is precisely the coercee's point of view which matters when it comes to understanding the unfreedom of her action, as opposed to merely the wrongfulness of the coercer's treatment of her. If everything is in order as far as the coercee's actual choice is concerned, then it is difficult to understand the sense in which the coercer has in fact subjected her will to his own.

[13] Compare Buss, "Valuing Autonomy and Respecting Persons," 226ff.

doing so. (Assume, for the sake of argument, that she is guilty of some fireable offense.)

To deal with the threat of extensional inadequacy, moralists typically redescribe the threatened actions so that they turn out to be independently wrong after all. For example, Alan Wertheimer suggests that a blackmailer's threat might be wrong because "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."[14] Japa Pallikkathayil suggests that *quid pro quo* sexual harassment might be wrong because employment-at-will is independently wrong.[15] And T.M. Scanlon suggests that the action which the employer threatens is not simply *firing the employee* but *firing her because she did not comply with his threat*.[16] This more specific action, he adds helpfully, would be wrong 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," a point which Scanlon takes to apply to "abuses of privilege" in general.[17]

But this last suggestion shows what is bound to remain unsatisfactory about any moralistic treatment of the counterexamples. Scanlon makes a plausible point. If we ask *why* it is wrong to abuse one's power or privilege in the particular ways that characterize blackmail and *quid pro quo* sexual harassment, the most obvious answer is that when the power or privilege is weaponized in these ways, it constitutes an objectionable form of control by one person of another's conduct. Now, presumably not all ways of influencing other people's conduct count as objectionable control; ordinary rational persuasion, for example, does not. On what grounds, then, can we say that the kind of control which is exercised when power is weaponized, as in blackmail or *quid pro quo* sexual harassment, is distinctively objectionable? The obvious answer is that this is a kind of control which distinctively compromises the other person's self-determination; it is a way of subjecting another person's will to one's own or using them as a mere means. But moralistic theories are not entitled to this obvious answer, for the whole point of these theories is meant to be that the

[14] Wertheimer, *Coercion*, 220.
[15] 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, and that "acting on the employer's conditional intention involves paying for sex," which "might be impermissible."
[16] Scanlon, *Moral Dimensions*, 84.
[17] *Ibid.*, 84, 86.

wrongfulness of coercion can be understood entirely on other, independent grounds.

The fundamental problem, then, is not merely a matter of extensional inadequacy. It concerns the *structure* of the proposed explanation. Moralistic theories are not entitled to the most plausible form of explanation of the wrongfulness of acts like blackmail and *quid pro quo* sexual harassment, namely that such acts are wrongful *because* they are coercive, that is, because they are essentially ways of compromising another person's freedom. And this point is really a completely general one, not just about blackmail and sexual harassment but about coercive acts as such. Such acts are normally wrong because they compromise another person's freedom. In order to understand the moral significance of coercion, then, we cannot simply put aside the conceptual problem why coerced action is unfree.

C. Psychologism

Philosophers working in a very different, empiricist tradition have approached the problem from a seemingly different angle. What we might call psychologistic theories—defended by John Plamenatz, Gerald Dworkin, Harry Frankfurt, and many others—make a direct attempt to address the problem of unfreedom. The key is supposed to lie within the psychology of the coerced agent—more particularly, in the agent's reflective attitudes concerning the way her action is motivated. When we do something in order to avoid a threatened punishment, we do what we want in the sense that we do what we intend. But as Dworkin puts it,

> it is the attitude a man takes toward the reasons for which he acts, whether or not he identifies himself with these reasons, assimilates them to himself, which is crucial for determining whether or not he acts freely. Men resent acting for certain reasons; they would not choose to be motivated in certain ways.[18]

In particular, we tend to resent being motivated by threats. Frankfurt suggests:

> a person who submits to […] 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.[19]

---

[18] Gerald Dworkin, "Acting Freely," *Noûs* 4, no. 4 (1970): 377.

[19] Harry Frankfurt, "Coercion and Moral Responsibility," in *Essays on Freedom of Action*, ed. Ted Honderich (London: Routledge & Kegan Paul, 1973), 82–83.

For example, the mugging victim who hands over her wallet to save her life can be expected to resent the way she is motivated because we resent having to do something just to avoid being shot, or more generally just to prevent our situation from becoming worse. When we are coerced, then, there is a sense in which we are alienated from our motive. For this reason, a coerced action is something less than an unqualified, wholehearted expression of the agent's own will. She does it less than fully freely. And once again, it is commonly thought that we can explain the difference between threats and offers on the same grounds. For we do not typically resent doing something for the sake of obtaining an offered reward. This psychological difference, accordingly, can explain what is distinctively wrong with coercion. Or so it is argued.[20]

Psychologistic theories are in trouble if the relevant kind of resentment (or comparable reflective attitude) is understood simply as a brute psychological fact about the agent. For actual psychological resentment and unfreedom do not necessarily go together. Our actual dispositions to feel resentment can be quite idiosyncratic and arbitrary. On the one hand, people can in principle resent what does not make them unfree, perhaps if they have an exaggerated and unreasonable sense of entitlement. On the other hand, more problematically, people can in principle fail to resent what actually makes them unfree. A commonly cited example is the fanciful but nonetheless conceivable case of the "contented slave," who does not mind being coerced by her master, for

---

Compare, e.g., J. P. Plamenatz, *Consent, Freedom and Political Obligation* (London: Oxford University Press, 1938), 125; Dworkin, "Acting Freely," 377; Irving Thalberg, "Hierarchical Analyses of Unfree Action," *Canadian Journal of Philosophy* 8, no. 2 (1978); James Stacey Taylor, "Autonomy, Duress, and Coercion," *Social Philosophy and Policy* 20, no. 2 (2003); John Christman, "Introduction," in *The Inner Citadel: Essays on Individual Autonomy* (Brattleboro: Echo Point Books & Media, 2014); and significant parts of Suzy Killmister, *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance* (Routledge, 2018).

[20] Arguably, this conclusion is in tension with empirical research suggesting that people resent being offered rewards when they perceive these rewards as attempts to control their conduct and hence as attacks on their autonomy. See, e.g., 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). I am more interested, though, in challenging psychologism on grounds of principle than on empirical grounds.

reasons that are not attributable to impaired rationality, ignorance, or the like.[21] When this person is forced to do something by her master, she surely acts unfreely. But since *ex hypothesi* she does not resent the way she is motivated, psychologistic theories seem unable to affirm this obvious truth. These theories thus threaten to destroy the distinction between feeling unfree and actually being unfree.

For similar reasons, if the relevant reflective attitudes are brute psychological facts, it is unclear how psychologistic theories can explain the distinctive moral significance of coercion. We take it that we generally have at least defeasible second-personal claims against other people not to coerce us. But if the unfreedom of coerced action is at bottom merely a matter of some attitude of the agent toward it, it is not clear that it could give rise to such claims. We might not desire to be treated in this way, but it is doubtful that we have a general claim not to be subjected to any treatment which, as a matter of our contingent subjective make-up, we happen to resent.

In light of these concerns, it seems more promising to think of the relevant kind of resentment not as a brute psychological fact but as an essentially reasons-responsive attitude. Resentment is not merely a reaction to a stimulus; it is capable of being warranted by its object. In particular, it is clearly warranted by coercive threats. So perhaps the point is that coerced action is unfree because it tends to give rise to *reasonable* resentment on the part of the coerced agent. And reasonable resentment is neither arbitrary nor morally insignificant. Thus, Robert Nozick argues in his early article on coercion that the reason coerced action is distinctively non-voluntary is that "the Rational Man" feels an aversion at being threatened with punishment but not at being offered a reward.[22]

But in fact, this more rationalist line of thought reveals a deep and surprising similarity between psychologistic and moralistic theories. For thinking of resentment as a reasons-responsive attitude immediately invites the question *why* coercive threats reasonably give rise to resentment, or why Nozick's Rational Man feels an aversion at being threatened. The most natural and obvious answer is surely that coercive threats warrant resentment because they compromise one's self-determination. But this answer is incompatible

[21] For more detailed discussions of this sort of case, see, e.g., Marina Oshana, *Personal Autonomy in Society* (Aldershot: Ashgate Publishing, 2006), Chapter 3; and Killmister, *Taking the Measure of Autonomy*, Chapter 6.

[22] Robert Nozick, "Coercion," in *Philosophy, Science, and Method*, ed. S. Morgenbesser, P. Suppes, and M. White (New York: 1969), 460. I will put aside familiar worries about hypothetical consent here.

with the structure of the psychologistic explanation of the unfreedom of coerced action, which makes the resentment prior to the unfreedom. As Dworkin makes explicit, from a psychologistic point of view "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."[23] In order to explain why we find it "painful" to act for these reasons, then, the psychologistic theory must look to some other objectionable feature of coercive threats which gives rise to resentment. And this means it will not be able to explain what is objectionable about coercion—why coercion warrants resentment—in the most natural way, as fundamentally a function of the unfreedom of coerced action.

In other words, we are back to the problem with moralistic theories. Like moralistic theories, psychologistic theories are unable to explain what is wrong with coercion in the right way. They cannot put freedom first but must make the wrongfulness of coercion parasitic on some other, independent wrong. In a way, these two types of theory are just objective and subjective faces of one and the same idea. From a psychologistic point of view, the unfreedom of coerced action is parasitic on the psychological response to being coerced. But once we understand that this psychological response is a warranted response to being wronged, we can see that the proposed explanation is no different, fundamentally, from the moralistic one.

III. A REINTERPRETATION OF TRADITIONAL KANTIANISM

To understand coercion, we must put freedom first. As we saw earlier, however, this turns out to be difficult. For to understand the unfreedom of coerced action, we must solve the problem of the inalienability of agency, which the traditional Kantians seemed unable to solve. What these Kantians seemed unable to explain is the possibility of an action that is more than a mere passive movement on the one hand yet less than a full, unqualified exercise of practical reason on the other. What we need, then, is an account of the distinctive sense in which a coerced action, though intentional, is not underwritten in a full, unqualified way by the agent's own practical reason. In other words, we need an account of the sense in which the coerced agent, insofar as she is coerced, does not without qualification take what she is doing to be worth doing—she does not without qualification take herself to have good reason to do it—even as she intentionally chooses to do it anyway.

[23] Dworkin, "Acting Freely," 378–79.

In order to make progress toward such an account, I want to suggest taking a step back and examining the broader genus of interpersonal influence to which coercive threats belong and which distinguishes them most fundamentally from ordinary, autonomy-preserving rational persuasion. This is the genus of *incentivization*. The category of incentivization comprises not only threats of punishment but also offers of reward. In the literature, offers of reward have tended to figure only by way of *contrast* with threats of punishment, since offers, unlike threats, have commonly been assumed to be autonomy-preserving and morally innocuous as such. I will argue, however, that it is only once we have properly gotten into view what offers and threats have in common that we will be in a position to correctly identify what makes coercion by means of threats distinctively problematic.

A. Threats and Offers as Species of Incentivization

Both threats and offers of the relevant kind are species of incentivization. Incentivization can be defined as an attempt to get another person to do something by means of a prospective good which is conditional on but extrinsic to the relevant action. Let me explain.

If I want you to do something, one straightforward way I can try to get you to do it is by pointing out that the relevant action would constitute or produce some good which I know you care about. This is how rational persuasion of the most ordinary kind works. For example, suppose I am about to have to drive to a job interview but my car has broken down; and suppose you are a budding amateur mechanic. I might try to persuade you to fix my car on the grounds that this would enable me to get to my interview on time, or on the grounds that you might enjoy the challenge of practicing your skills, or both. If I succeed, you will fix my car, and your motive will be the benefit of enabling me to get to where I need to be, or the enjoyment inherent in practicing your skills, or both.

Another way I could try to get you to do what I want is by incentivizing you: by threatening you or making you an offer. I could threaten to beat you up unless you fix my car. Or I could offer you money for fixing my car. From one point of view, incentivization of either sort looks just like the case of ordinary rational persuasion. If I succeed, you will end up fixing my car, and you will do so for the sake of some prospective good: avoiding a beating, getting some money.

There is, however, this crucial difference. In the case of ordinary rational persuasion, the good for the sake of which I get you to act is one to which the activity already makes some productive or constitutive contribution,

independently of my attempt to get you to do it. My intervention was not needed to make it the case that fixing my car would enable me to get to work on time, or that it would enable you to practice your skills. When I persuade you to do it for either of these reasons, I merely draw your attention to a benefit which the activity already brings about or involves. By contrast, when I incentivize you, the good for the sake of which I get you to act is not one to which the activity itself already makes a contribution, independently of my attempt to get you to do it. When I persuade you to fix my car for the sake of avoiding a beating or getting some money, I do not merely draw your attention to an existing benefit of this activity. Rather, I go out of my way to attach the benefit to the action.

We can express the difference by saying that the reason for which you act in the case of ordinary rational persuasion is *intrinsic* to your action, and your action is intrinsically motivated, whereas in the case of incentivization your reason is *extrinsic* to your action, and the action is extrinsically motivated. This is a different intrinsic–extrinsic distinction than the more familiar one between activity for its own sake and activity for the sake of a further end. When you fix my car to enable me to get to my job interview, you act for the sake of a further end, a product which is distinct from the activity that brings it about. But that end, though distinct from the activity, can nonetheless be considered internally connected with the activity in the sense that the activity itself already makes some contribution to your end. By contrast, when you fix my car to avoid a beating or to get some money, there is a sense in which the activity of fixing my car itself makes no contribution to your end. Your end consists *neither* in some aspect of the activity itself *nor* in a product of the activity distinct from it; it really has nothing to do with the activity at all.

It is essential to incentives, given their function, that they are in this sense extrinsic to what they incentivize. For if the action itself already produced a good sufficient to motivate you to do it, then there would be no need for me to add a stick or carrot. The action would come, as it were, prepackaged with its own carrot. In general, when I incentivize you to do something by threatening you with punishment or offering you a reward, I do so precisely on the assumption that the thing I want you to do—the activity itself—might neither constitute nor contribute to a good which is already sufficient to motivate you to do it. That is why I feel the need to give you an incentive in the first place.

The intrinsic–extrinsic distinction I am drawing is somewhat subtle; it requires us to distinguish between two different ways of acting for the sake of a further end that we often assimilate to each other. Without some version of this distinction, however, it is hard to see how we could adequately explain the

undeniable difference between incentivization and other forms of interpersonal influence. And indeed the distinction is not theoretically novel but is familiar from the philosophical literature on incentives. Ruth Grant, for instance, defines an incentive as, among other things, "an extrinsic benefit," that is, one which is not "the natural or automatic consequence of an action."[24] She also makes the observation that "[i]f the desired action would result naturally or automatically [meaning, presumably, in view of the expected natural or automatic consequences of the action alone], no incentive would be necessary. An incentive is the added element without which the desired action probably would not occur."[25]

B. Incentivization as Heteronomy

Let us go back now to the difficulty with traditional Kantianism. To reiterate, the problem was as follows. We want to say that a volitionally coerced agent does not act on the basis of her own judgment that what she is doing is worth doing—that her practical reason is in some way supplanted or subverted by her coercer. But at the same time, we cannot plausibly deny that even this coerced agent, inasmuch as she acts intentionally, exercises her own practical reason and pursues her own ends in acting as she does. The only way out of the difficulty, it seems, is to articulate a sense in which she exercises her practical reason only in a privative, qualified way. But the traditional Kantians have no account of what this could mean. In what sense can someone be acting intentionally and yet fail to be choosing their action on the basis of their own unqualified judgment of what they have reason to do?

The intrinsic–extrinsic distinction can help us make sense of such a privative exercise of practical reason. Insofar as you act for an intrinsic reason, you choose your action on its own merits, that is, on the basis of considerations that speak in favor of the action itself, independently of my attempt to get you to do it. Insofar as you act for an extrinsic reason, on the other hand, you do not choose your action on its own merits. But this seems to be precisely the sense in which an agent acting under coercion is not motivated by her unqualified judgment that there is good reason to do what she does. To be sure, she acts intentionally and so judges her action to be worth doing in some thin

[24] Ruth W. Grant, *Strings Attached: Untangling the Ethics of Incentives* (Princeton: Princeton University Press, 2012), 43–44. In Grant's usage, threats do not count as incentives, but this makes no relevant difference.
[25] Grant, *Strings Attached*, 43–44.

sense—but this judgment is essentially only a qualified endorsement, for it is not based on her consideration of *the action* on its own merits.

This is easiest to see in those cases where someone acts for the sake of an incentive and does not think that *anything* else speaks in favor of her action. Cases of coercion are perhaps typically like this, but so are some cases of incentivization by means of offers. Consider the well-known phenomenon of "bullshit jobs": jobs which are such that the people doing them take them to be pointless.[26] The idea is not merely that such jobs are undesirable in themselves. Someone might perform a very boring and unpleasant job which they nonetheless think is very much worth doing because it makes an important social contribution; this would not be a bullshit job. A bullshit job is one which is such that the person doing the job takes it to be valuable *neither* in itself *nor* on account of its product. Now, someone who intentionally performs such a job in order to pay her bills clearly takes herself to have reason to do it in some thin sense. But it is equally clear that there is also a more demanding sense in which she does not take herself to have reason to do it. And on this basis, we can say plausibly that her activity is not underwritten in a full and unqualified way by her own practical reason. But on traditional Kantian grounds, it follows that someone motivated in this way does not act fully autonomously.

Because extrinsically motivated action is less than fully autonomous, to attempt to motivate another person in this way is normally *pro tanto* wrong. It is wrong on familiar Kantian grounds: because it compromises the other's autonomy, failing to respect them fully as a rational agent. When I use ordinary rational persuasion to get you to do something, as Korsgaard says, I put the facts before you and ask for your contribution.[27] In doing so, I allow you to choose your action on its own merits, independently of my attempt to get you to do it. I thereby treat you as an end in yourself because I respect your status as an independent arbiter of what you have reason to do. I refuse to supplant your practical reason by my own. This is risky, of course, since you might not come to the desired conclusion when faced with the relevant facts. But that risk is an unavoidable hazard for rational agents who depend on one another but are at the same time committed to respecting one another's independent agency.

When, on the other hand, I incentivize you to do something, I am not prepared to rely on your judgment as to whether the action is worth doing on its own merits. If my attempt to influence you succeeds, then your action will

[26] See David Graeber, *Bullshit Jobs: A Theory* (New York: Simon & Schuster, 2018).
[27] Korsgaard, "The Right to Lie," 142.

not be autonomous; it will not have its source in your own unqualified judgment that this is what you have reason to do. In this way, I fail to respect your status as an independent arbiter of what you have reason to do. I use you as a mere means, in the sense that I use your power of practical reason to get you to do something not fully underwritten by your own exercise of this power. My attitude toward you rather resembles the attitude which Korsgaard says a deceiver takes toward his victim. In deception, she says,

> Your reason is worked, like a machine: the deceiver tries to determine what levers to pull to get the desired results from you. Physical coercion treats someone's person as a tool; lying treats someone's *reason* as a tool. This is why Kant finds it so horrifying; it is a direct violation of autonomy.[28]

This description could just as well have been written about incentivization rather than deception. Incentivization, like deception, is heteronomy.

To be sure, it can be hard to see how the mere act of making someone an offer, *considered on its own*, could count as compromising their autonomy. After all, one merely gives the other an option which they would not otherwise have had, and one does not literally force them to accept the offer, which they are at liberty to "take or leave." In order fully to understand the nature of offers, however, it is not enough to focus narrowly on the act of making an offer by itself; we must also attend to what this act is meant to achieve. While I have sometimes spoken as though simply making an offer to someone were sufficient to compromise their autonomy, this is just a convenient shorthand. More precisely, my claim is that to make someone an offer is to *attempt* to do something which will compromise the other's autonomy to the extent that it succeeds. An offer of the relevant kind is an attempt to get someone to perform a certain action, and to the extent that it succeeds, the other will be motivated to do that thing by an extrinsic reason. To be motivated in this way, I have argued, is to act unfreely. To make a *successful* offer is accordingly to succeed in getting someone to act unfreely.

The need to broaden our focus in this way is not peculiar to offers; something similar, it seems to me, is true of coercion. When I threaten you in the relevant sense, I attempt to get you to do something. If I succeed, you will do that thing for an extrinsic reason, as a way of avoiding the threatened punishment. I argued that to be motivated in this way is to act unfreely. Therefore, to make a *successful* threat is to succeed in getting someone to act unfreely. It is a somewhat different story if the attempt fails. Suppose I threaten you with a beating unless you do something for me—but instead of doing the thing, you

[28] *Ibid.*, 140–41.

refuse. I can now try to beat you and perhaps I will succeed in doing that, but even so, I have not succeeded in subjecting your will to mine as far as the action which I wanted you to perform is concerned. Your compliance with my command would have been an unfree choice, but your refusal is a free choice.[29] This seems quite plausible. As in the case of offers, then, in order fully to understand why coercion is a way of making another person unfree, we must look beyond the act of issuing a threat by itself and attend to its purpose—attend, in other words, to what it would be like for the attempt at influencing another's behavior in this way to succeed.

We can say, then, that incentivization is wrong because it is an attempt to do something which will make someone unfree if it succeeds. And what is true of incentivization in general is true of coercion in particular. Coercion is wrong because coerced action, as a species of extrinsically motivated activity, is a privative exercise of practical reason and less than fully autonomous. The moral foundation of this explanation is each person's right to freedom—each person's claim to be treated by others only in ways consistent with her own autonomy. A defense of this very basic right is beyond the scope of this paper. But grounding the wrongfulness of coercion in such a right genuinely puts freedom first and thus gives our theory of coercion the correct normative structure at last.

There is a widespread assumption in the literature that offers as such, unlike threats, are autonomy-preserving and morally innocuous. My argument to this point amounts to a principled criticism of this assumption. If I am right, then the assumption stands in the way of an adequate theory of coercion and must be abandoned. We can only understand the unfreedom of coerced action and the wrongfulness of coercion as a species of the unfreedom of extrinsically motivated activity in general, and the wrongfulness of making another unfree in this way.

This is not to claim that threats and offers are morally equivalent. As I will go on to argue shortly, there is indeed a morally significant difference between coercive threats and offers of reward. But the point is that whatever this difference turns out to be, it is not as black and white as has been assumed. It is not the case that coercive threats compromise autonomy while offers of

[29] This is what Hegel has in mind when he says provocatively, "Only he who *wills* to be *coerced* can be coerced into anything." G. W. F. Hegel, *Elements of the Philosophy of Right*, trans. H. B. Nisbet, ed. Allen W. Wood (Cambridge: Cambridge University Press, 1991), §91; compare the Remark and Addition to §57, and the Addition to §99. Of course, my point in this paper has been that this cannot be the whole story (as Hegel also recognizes).

reward leave it intact. Nor is it the case that coercive threats are wrong while offers of reward are morally innocuous. Both threats and offers are attempts to exercise a kind of control over another's conduct that is incompatible with the other's autonomy; and if there is a right to freedom, then both are normally *pro tanto* wrong for that reason.

C. What Makes Coercion Special

Even if all incentivization is morally objectionable in some way, there still seems to be a moral difference between threats and offers, and a theory of coercion would not be complete without an account of this difference. While my argument so far does not depend on any particular such account, I will now sketch what I take to be the most promising line of thought. In this way, I show that taking a more critical attitude toward incentivization as such does not preclude us from doing justice to what makes coercion special as compared with other species of incentivization.

When one person, *A*, incentivizes another, *B*, to do something, *A* attempts to exercise control over *B*'s conduct. This control depends on *A* having and exercising some measure of power over *B*'s access to some good.[30] Since this power makes *B* in some measure vulnerable to *A* in respect of access to the relevant good, we can say, by the same token, that incentivization is essentially a way of taking advantage of someone's vulnerability.[31] The difference between threats and offers, I will suggest, consists in the extent to which *A*'s power over *B*'s access to the relevant good is unilateral, or—what amounts to

[30] Note that for the purpose of this paper I am simply putting aside cases of deceptive incentivization, where someone makes a threat or an offer which they are unable or unwilling to follow through on—in other words, where they are bluffing. An account of these cases will likely have to be parasitic on an account of truthful incentivization (and perhaps additionally on an account of deception).

[31] Though I do not insist on the terminology here, I think this makes it reasonable to classify incentivization as exploitation. On one classical understanding of exploitation, to exploit someone is to take advantage of their vulnerability in a way which wrongs them. (For a view along these lines, see, e.g., Robert E. Goodin, "Exploiting a Situation and Exploiting a Person," in *Modern Theories of Exploitation*, ed. Andrew Reeve (London: Sage, 1987).) But *A*, in incentivizing *B*, takes advantage of *B*'s vulnerability to *A*—in respect of *B*'s access to some good—in order to get *B* to act in a way which would necessarily be unfree, thus violating *B*'s right to freedom. In other words, *A* exploits *B*.

the same thing—the extent to which *B*'s vulnerability to *A* is a product of *A*'s will alone.

Let us analyze the relevant forms of power and vulnerability in more detail. When *B* acts under the influence of *A*'s incentive, she chooses her action instrumentally, as a means to the good which *A* controls. This presupposes, first, that *A* is able to help make it the case that *B* will in fact get the good she is after if she complies. In other words, *A* must have the power to make it the case that the incentivized action is a possible means to *B*'s end. For example, if *A* offers *B* money in exchange for *B*'s labor, the premise of the transaction is that *A* has the power to make it the case that *B* will in fact acquire the money in the event that she performs the labor, thus turning the labor into a possible means to acquiring the money.

But in general, a successful act of incentivization presupposes more than this. It must be the case not only that the incentivized action is *a* means to *B*'s end but also, second, that it is a more desirable means than the available alternatives that *B* considers. For example, if there is a $100 bill lying on the ground in between *A* and *B*, and *A* offers to pick it up and give it to *B* in exchange for one day's hard labor, *A* thereby makes it the case that performing the labor is *a* possible means to getting the $100. However, this attempt at incentivization is hardly likely to succeed, since *B* has available to her the far more desirable means of stooping and picking the money up herself, without having to perform any hard labor for *A*. For *A*'s attempt to succeed, it would have to be the case that *B* judges the incentivized action to be a more desirable means to her end than the available alternative means.

Let us see how this works in the case of threats. First, as in the case of offers, *A* must be able to make it the case that the incentivized action is a possible means to *B*'s end. For example, if *A* threatens to shoot *B* unless *B* hands over her money, the premise of this transaction is that *A* has the power *not* to shoot *B* in the event that *B* complies, thereby making this compliance a possible means to staying alive. If it were clear that *A* was going to shoot *B* regardless, or if *B* were anyway certain to die imminently for independent reasons, compliance would no longer be a means to staying alive and the incentive would fail to have its desired effect. Second, as in the case of offers, the incentivized action must be more desirable than the available alternatives which *B* considers. Of course, when *A* threatens to shoot *B* unless *B* complies, the point is that there are no longer *any* alternative means to *B*'s end of survival, for if *B* does anything but comply, *A* will shoot *B* dead. Thus, the incentivized action trivially becomes the most desirable means to *B*'s end.

The key difference between threats and offers concerns the lengths to which *A* is prepared to go in securing this second condition. When *A* makes

B a pure offer, *A* unilaterally—by his own will alone—thereby intends to make *B*'s incentivized action a possible means to *B*'s end, but *A* does not thereby intend to go out of his way to make other possible means to the same end absolutely less desirable or altogether unavailable to *B*. For this reason, *A* allows the relative desirability of the incentivized action as a means to the relevant end, and hence the success of *A*'s attempt of incentivization, to depend essentially on facts beyond *A*'s own will, including the actions of other people as well as natural circumstances and random chance. The control which he intentionally exercises over *B*'s access to the relevant good, and thereby over *B*'s conduct, is to that extent not unilateral but only partial.

For example, if *A* offers *B* money in exchange for labor, *A* does not thereby undertake to block other possible routes of access to money in the event that *B* refuses the offer. He does not, for instance, intentionally prevent *B* from coming into an unexpected inheritance, or intentionally prevent other would-be employers from presenting *B* with competing offers. Now, it may *in fact* be the case that *B* finds all of the alternative ways of getting the money less desirable than working for *A*. Indeed, it may in fact be the case that *B* has no alternative ways to get the money at all. For example, perhaps the prevailing economic system makes it impossible for her to access money except by exchanging her labor for it, and perhaps *A* happens to be the only employer hiring in *B*'s area. Insofar as we are dealing with a pure case of an offer, however, this fact can only be—from *A*'s point of view—a happy coincidence. It is not itself a product of *A*'s will, but merely a circumstance of which he takes advantage for the purpose of *B*'s conduct.

By contrast, when *A* coerces *B* by means of a threat, *A* is not prepared to leave it to chance or to other people whether *B* will have more desirable means to her end. Rather, he undertakes to go out of his way to block all such alternative means. He thus unilaterally—by his own will alone—intends to make it the case not only that *B*'s incentivized action is a possible means to *B*'s end but that it is the only possible means to *B*'s end, a *sine qua non* of it. In this way, *A* takes a distinctively unilateral and distinctively secure control of *B*'s access to the relevant good, and thereby over *B*'s conduct.

For example, if *A* threatens to shoot *B* unless *B* hands over her money, *A* undertakes to make it the case by his will alone that *B* shall not have access to the good she is after—her survival—unless she complies with his demand. In the event that she refuses to comply, he is not prepared to leave the matter of her survival to chance or to the will of other people. Rather, he intends to

take matters entirely into his own hands and singlehandedly make it the case that she will not survive.[32]

To be sure, since no one is omnipotent, *A*'s ability to carry out this intention is bound to depend on facts outside *A*'s individual control, including the willingness and ability of other agents to intervene or to support his behavior.[33] But the point remains that, given whatever *de facto* power *A* has to block alternative means to his victim's end, *A*, in issuing a threat rather than an offer, fully intends to exercise this power. When *A* makes *B* an offer, he may well happen to have similar *de facto* power to block some or all of *B*'s alternative options, but the point is that he does not, in making the offer, intentionally exercise this power. What distinguishes coercion by means of threat from other forms of incentivization, then, is that a coercer intentionally exercises a distinctively unilateral form of control over his victim.

What is the moral significance of this difference? Previously, I argued that both threats and offers of the relevant kind are attempts to get another person to do something, in a way which is incompatible with the other's autonomy. One who makes an offer, however, leaves more to chance and to the will of other people than one who makes a threat. Thus, one who makes a threat displays an especially uncompromising commitment to subjecting his victim's will to his own. From a moral point of view, then, we can say that an

---

[32] 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 threats from offers. The kind of "tracking" on which I am focusing, on the other hand, seems more essential to coercive threats, and also distinguishes them from offers: 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.

[33] 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). on the juridical context which enables the coercive relation between master and slave.

offer is related to a threat in something akin to the way in which opportunistic wrongdoing is related to premeditated wrongdoing (though the relation is not a specifically temporal one in this case). Both threats and offers are ways of taking advantage of another person's vulnerability—her desire for a good over which one has some control—in order to control that person's conduct. But one who makes an offer does not go out of his way to exacerbate that vulnerability; he *merely* takes advantage of it. One who makes a threat, on the other hand, intentionally makes the other person more vulnerable to him, by intentionally depriving her of other means of accessing the relevant good, for the purpose of controlling her conduct. For this reason, he wrongs her in a more serious way, more flagrantly violating her claim to autonomy.

I take this proposal to be fairly commonsensical. But it depends essentially on my foregoing criticism of incentivization as a genus. For it depends on a prior understanding of what is wrong with attempting to control another's conduct by controlling their access to some good in the first place. Let us recall the difficulty. Even if a mugging victim has no way of achieving her end of staying alive other than to hand over her money to the mugger, it remains the case that she is the one who chooses her end of staying alive and that she is the one who chooses her means to this end, given her circumstances. Her action is a genuine exercise of her own practical reason. We were able to understand why it is a privative exercise of practical reason only by recognizing that extrinsically motivated action *in general* is less than fully underwritten by the agent's own practical reason. So the distinction between threats and offers, while morally significant, remains a distinction within heteronomy.

IV. AUTONOMY AND THE MARKET

When we talk of offers, thoughts of the market are never far away. In order to clarify and refine my view, I will close by outlining some of its implications for the moral standing of the market. These implications are nuanced, to some extent underdetermined, and in need of much further elaboration which is beyond the scope of this paper. Broadly speaking, though, my argument about autonomy and coercion suggests that the market as we know it, while an improvement on straightforwardly coercive alternatives, is incompatible with full individual autonomy.[34]

[34] Note that my discussion will abstract from the distinction between labor markets and other sorts of market, and hence also from the distinction

A. Freedom from Unilateral Coercion

My account of the difference between threats and offers has certain pro-market implications, at least in theory.

Importantly, the account captures a traditional and sensible understanding of what favorably distinguishes social relations in a market from those between master and slave and between lord and serf. A master controls his slave's conduct in an intentionally unilateral way: when he threatens to deprive his slave of dinner unless she does what he wants, he intentionally ensures that the *only* way she can get her dinner is by obeying him. Something similar is true of the relation between lord and serf. By contrast, at least in theory, neither participant in a market transaction exercises such unilateral power over the other. If I pay you to fix my car, this exchange may in fact be what enables you to get your dinner. But insofar as I do not prevent you from getting the money you need in some other way, for example by preventing you from contracting with some other customer instead, I do not intentionally subject your conduct to my own will alone. Likewise, insofar as you do not prevent me from getting my car fixed in some other way, for example by preventing me from contracting with another mechanic instead, you do not intentionally subject my conduct to your will alone. This is the great leveling power of markets. As Adam Smith puts it:

> Each tradesman or artificer derives his subsistence from the employment, not of one, but of a hundred or a thousand different customers. Though in some measure obliged to them all, therefore, he is not absolutely dependent on any one of them.[35]

Now, of course it may happen, for example, that there is a dearth of mechanics in our town—so that I really have no choice but to contract with you if I am to get my car fixed—or a dearth of cars in need of repair, so that you must contract with me if you are to get your dinner. But even so, at least in theory, our lack of alternatives in either scenario is not a product of the other's will. The power which one of us exercises over the other is not intentionally unilateral. For example, in the event of a shortage of mechanics, you are not predisposed to prevent other mechanics from stepping into the breach and correcting the disequilibrium in the market. In this way, agents meeting in

---

between capitalists and workers. These distinctions raise large issues—too large for this paper.

[35] Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, ed. Edwin Cannan (Chicago: University of Chicago Press, 1976), 438.

the marketplace may take advantage of one another's incidental vulnerability but they do not seek the unilateral control characteristic of masters and lords—at least in theory.

Turning from theory to reality, this pro-market judgment requires some qualification. Even if we put aside blatant coercion involving violence, extortion, and the like, market transactions can in practice exhibit definite elements of coercion. This is intelligible because the unilaterality of the control which a person intentionally exercises over another's good is a matter of degree. For instance, if you offer to fix my car in exchange for money but you also go out of your way to monopolize the car repair market by actively preventing other mechanics from setting up shop, you intentionally exercise a more unilateral form of control over me. It is not fully unilateral. For example, you do not prevent me from trying to fix the car myself, or from finding a friend who will fix it for me. But the more you are prepared to go out of your way to prevent me from taking alternative means to my end, the more your "offer" takes on a coercive character.

That said, at least some of these problems could be prevented by ensuring that markets remain competitive, as they are theoretically supposed to be. Besides, other things equal, even partially coercive offers are preferable to full-fledged coercion. My account of the difference between threats and offers therefore gives us strong reason to prefer markets to some of their historical alternatives.

B. Reciprocal Heteronomy

If we set our sights higher and ask not merely whether market exchange is preferable to coercion but whether it is compatible with full-fledged individual freedom, the picture looks decidedly less rosy—though even here there will be nuances and caveats.

I have argued that incentivization in general compromises autonomy. Any more concrete conclusions we draw about the moral standing of the market will depend on the role of incentivization in the market. This in turn is bound to be a matter of degree, for two reasons. First, since a market comprises many individual transactions, some of these might in principle involve incentivization while others do not, so we will have to consider how many of them, relatively speaking, are likely to be motivated in which way. This will give us a sense of the extent to which individual unfreedom is implicated in the market as a whole social institution.

Second, it seems possible in principle to do something for the sake of a mix of intrinsic and extrinsic reasons. For example, many people, if their jobs

are not bullshit jobs, would likely claim to do the work they are paid to do in part for the money, an extrinsic end, but also in part for the sake of ends intrinsic to the work, such as the development and exercise of skills and the valuable social contribution which the work makes by virtue of its product. If we take this self-understanding at face value, it seems reasonable to say that such doubly motivated work is partly free and partly unfree, or free under one aspect and unfree under another. It is free insofar as it is motivated by intrinsic reasons, and unfree insofar as it is motivated by extrinsic reasons. Moreover, if it makes sense to say, for instance, that the one or the other motive is predominant, we can likewise say that the aspect of freedom or the aspect of unfreedom predominates.

This is a principled line which it would be equally reasonable to take with respect to coercion. Someone who is commanded, on pain of punishment, to do something which she considers worth doing anyway might—conceivably, if improbably—do it for two reasons at once: the extrinsic reason of avoiding punishment, and the intrinsic reason that speaks in favor of the action on its own merits. It seems reasonable to say of such an agent's action that it is freely chosen insofar as she chooses it on its own merits, and unfree insofar as she chooses it for the extrinsic reason of avoiding punishment. And we might likewise want to say that the aspect of freedom or unfreedom is predominant insofar as the one or the other motive predominates.[36]

To what extent, then, are real-world market transactions likely to involve incentivization? It seems safe to say: to a very large extent. I would venture that most of the time when one person sells a good or service to another, the vendor provides their good or service primarily for the purpose of getting the other's money in return, and the buyer provides their money primarily for the purpose of getting the other's good in return. Each uses the object of the other's desire primarily as an incentive to motivate the other's performance, and each is in fact primarily motivated in their own performance by the incentive which the other provides.

Indeed, the power of markets, especially in the context of large, complex, and anonymous social networks, is supposed to rest in part on this motivational structure, the great advantage of which is that it presupposes no bond of solidarity between the transacting parties. If I bake your bread so that you fix my car, and *vice versa*, we can reliably succeed in satisfying one another's desires even if—in what we might call the purest case of market exchange—the prospect of satisfying the other's desire moves us not at all. Of course, since we are human beings with social sentiments, we do often care about

[36] Compare Nozick, "Coercion," 464.

satisfying others' desires, and many particular real-life market transactions may depart from the pure case. But something closely approximating the pure case is likely to remain the norm in a market economy as we know it.

Defenders of markets have long and proudly recognized this fact. Most famously, there is Smith's remark:

> 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.[37]

Smith's point is not exactly that the butcher, the brewer, and the baker necessarily have *selfish* motives. As is often pointed out, that may not be the case: they may be motivated by the unselfish end of providing for their families, for example. The point is that even if they have other-regarding ends, what motivates them, in a paradigm case of market exchange, is not the good of the person with whom they are immediately transacting; what motivates them, for one reason or another, is rather the good that this person can provide for them in return: the incentive.

Perhaps it is theoretically possible for market exchange to exist on a large scale without any element of incentivization.[38] But it does seem that transactions in a market as we know it normally involve this extrinsic motivational structure. And it follows that the market as we know it normally involves unfreedom. Interestingly, this unfreedom takes the form of a distinctively reciprocal heteronomy. Since *each* party in a typical market exchange incentivizes the other, each party compromises the other's autonomy and uses the other as a mere means.[39] They thus reciprocally wrong each other.

What follows from this moral judgment? Perhaps less than one might think. For one thing, it is a *pro tanto* judgment whose practical implications

---

[37] Smith, *The Wealth of Nations*, 18.

[38] See, e.g., Joseph H. Carens, *Equality, Moral Incentives, and the Market: An Essay in Utopian Politico-Economic Theory* (Chicago: University of Chicago Press, 1981). See also Julius, "The Possibility of Exchange."

[39] In this way, we can understand and vindicate, to some extent, the early Marx's seemingly hyperbolic claim that "[t]he social relationship in which I stand to you [in production for exchange], 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." Karl Marx, "On James Mill," in *Karl Marx: Selected Writings*, ed. David McLellan (Oxford: Oxford University Press, 2000), 130.

remain indeterminate in advance of further argument. I do think that there is something morally objectionable about market relations insofar as they involve incentivization. But for the purpose of this paper, I am prepared to allow what a full-blooded Kantian would deny, namely that this consideration might be outweighed by others. For example, all of the non-coercive alternatives to the market as we know it might be disastrously inefficient.

I insist only on two things. First, in light of the high priority which individual freedom should have in any reasonable moral and political philosophy, we have reason to *look* for such an alternative. Second, whether or not there is a non-disastrous alternative, we should be more clear-eyed about the moral *cost* of the market's undoubted virtues. Defenders of the market have traditionally extolled it on one or both of two grounds: efficiency and individual freedom. Well, production for the market as we know it may be efficient, at least for certain purposes—but it is not without qualification free.

As I have said, the moral judgment about the market which I have defended does not yet imply an all-things-considered judgment. I should add that it also does not yet imply a judgment of moral blameworthiness. In general, it is one thing to determine whether an action is wrong, and another to determine whether and to what extent the agent deserves to be condemned for it. It is possible that in certain contexts, individual agents may be partially or wholly excused for genuine wrongdoing and therefore not liable to much or any condemnation—which does not mean that everything is morally in order. For example, consider the fact that we live in a society in which the average person cannot reliably secure the goods that she needs in order to live well except by incentivizing other people to share access to those goods with her. In this context, we use our control over what other people need in order to get what we want from them in an autonomy-compromising way, just as those others do the same to us. I have argued that this is *pro tanto* wrongful conduct on the part of each of us. But it would be too quick to condemn individual people for "choosing" to partake in such relations rather than running the risk of serious deprivation or death. If there is an appropriate object of condemnation here, it is in the first instance the form of social organization which makes each person's ability to lead a good life conditional on her participation in *pro tanto* mutually wrongful social relations.

VI. CONCLUSION

In the first half of this paper, I showed that existing Kantian theories of coercion fail to explain both the unfreedom of coerced action and the wrongfulness

of coercion. Traditional, non-moralized Kantian theories are attractive on paper because they ground the wrongfulness of coercion plausibly in the way coercion compromises autonomy. But defenders of this approach have not given us a way of making sense of (volitionally) coerced agency, which is neither fully self-determined action on the one hand nor mere passive movement on the other. Moralistic theories try to do without a solution to this problem by making the wrongfulness of coercion parasitic on some other, independent wrong. But this structure of explanation jettisons the very feature of the original Kantian approach which made it so attractive in the first place: the fact that it sought to explain the wrongfulness of coercion, plausibly, on the basis of how coercion compromises autonomy. Psychologistic approaches try to explain the unfreedom of coerced action on the basis of the coerced agent's distinctive psychological reaction to being coerced. But once again, the structure of the explanation is wrong. For surely we resent being coerced because coercion wrongs us, and coercion wrongs us fundamentally because it makes us unfree.

In the second half of the paper, I showed that we can make progress by understanding coercion as a species of incentivization. Incentivized action in general, I argued, has a distinctively privative motivational structure. To do something for the sake of obtaining an offered reward or avoiding a threatened punishment is to be motivated by an end extrinsic to one's action, rather than to choose one's action on its own merits. But on familiar Kantian grounds, this is to act less than fully autonomously. It is not to choose one's action in a way underwritten without qualification by one's own practical reason. Both threats and offers therefore compromise autonomy—and are distinctively wrongful for that reason.

The heteronomous control which is involved in incentivization operates by way of controlling another's access to some good. This observation allowed me to bring into view what makes coercion more egregiously wrong than incentivization in general. The coercer goes out of her way to ensure unilaterally that his victim will be deprived of the relevant good unless she complies. This is an especially committal form of heteronomous control, by comparison with which offers are merely opportunistic. However, we should still reject the assumption that offers, as forms of incentivization, are completely benign. To the extent that we confront one another with either sticks or carrots, our social relations are not worthy of rational agents and we do not live in a society of free people.

REFERENCES

Bazargan, Saba. "Moral Coercion." *Philosophers' Imprint* 14, no. 11 (2014).

Buss, Sarah. "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints." *Ethics* 115 (2005).

Carens, Joseph H. *Equality, Moral Incentives, and the Market: An Essay in Utopian Politico-Economic Theory*. Chicago: University of Chicago Press, 1981.

Christman, John. "Introduction." In *The Inner Citadel: Essays on Individual Autonomy*, 3–23. Brattleboro: Echo Point Books & Media, 2014.

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): 627–68.

Dworkin, Gerald. "Acting Freely." *Noûs* 4, no. 4 (1970): 367–83.

Frankfurt, Harry. "Coercion and Moral Responsibility." In *Essays on Freedom of Action*, edited by Ted Honderich, 63–86. London: Routledge & Kegan Paul, 1973.

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, 155–79. Minneapolis: University of Minnesota Press, 2002.

Goodin, Robert E. "Exploiting a Situation and Exploiting a Person." In *Modern Theories of Exploitation*, edited by Andrew Reeve, 166–200. London: Sage, 1987.

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.

Hegel, G. W. F. *Elements of the Philosophy of Right*. Translated by H. B. Nisbet. Edited by Allen W. Wood. Cambridge: Cambridge University Press, 1991.

Herman, Barbara. "Leaving Deontology Behind." In *The Practice of Moral Judgment*, 208–40. Cambridge, MA: Harvard University Press, 1993.

Julius, A. J. "The Possibility of Exchange." *Politics, Philosophy & Economics* 12, no. 4 (2013): 361–74.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press, 1998.

Killmister, Suzy. *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance*. Routledge, 2018.

Korsgaard, Christine M. "Kant's Formula of Humanity." In *Creating the Kingdom of Ends*, 106–32. Cambridge: Cambridge University Press, 1996.

———. "The Right to Lie: Kant on Dealing with Evil." In *Creating the Kingdom of Ends*, 133–58. Cambridge: Cambridge University Press, 1996.

Marx, Karl. "On James Mill." In *Karl Marx: Selected Writings*, edited by David McLellan, 124–33. Oxford: Oxford University Press, 2000.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

———. "Coercion." In *Philosophy, Science, and Method*, edited by S. Morgenbesser, P. Suppes and M. White. New York, 1969.

O'Neill, Onora. "Between Consenting Adults." *Philosophy and Public Affairs* 14, no. 3 (1985): 252–77.

Oshana, Marina. *Personal Autonomy in Society*. Aldershot: Ashgate Publishing, 2006.

Pallikkathayil, Japa. "The Possibility of Choice: Three Accounts of the Problem with Coercion." *Philosophers' Imprint* 11, no. 16 (2011).

Plamenatz, J. P. *Consent, Freedom and Political Obligation*. London: Oxford University Press, 1938.

Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press, 2009.

Sachs, Benjamin. "Why Coercion Is Wrong When It's Wrong." *Australasian Journal of Philosophy* 91, no. 1 (2013): 63–82.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press, 2008.

Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Edited by Edwin Cannan. Chicago: University of Chicago Press, 1976.

Taylor, James Stacey. "Autonomy, Duress, and Coercion." *Social Philosophy and Policy* 20, no. 2 (2003): 127–55.

Thalberg, Irving. "Hierarchical Analyses of Unfree Action." *Canadian Journal of Philosophy* 8, no. 2 (1978): 211–26.

Vrousalis, Nicholas. "The Capitalist Cage: Structural Domination and Collective Agency in the Market." *Journal of Applied Philosophy* 38, no. 1 (2021): 40–54.

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.

Yaffe, Gideon. "Indoctrination, Coercion and Freedom of the Will." *Philosophy and Phenomenological Research* 67, no. 2 (2003): 335–56.

Zimmerman, David. "Coercive Wage Offers." *Philosophy and Public Affairs* 10, no. 2 (1981): 121–45.