Weakness of Will and the Measurement of Freedom*

Nicolas Côté

This article argues for a novel approach to the measurement of freedom of choice, on which the availability of an option is a matter of degree, rather than a bivalent matter of being either available or not. This approach is motivated by case studies involving weakness of will, where deficiencies in willpower seem to impair individual freedom by making certain alternatives much harder to pursue. This approach is perfectly general, however: its graded analysis of option availability can be extended to cases where weakness of will is not involved and can be used to generalize numerous existing measures of freedom.

I. INTRODUCTION

Weakness of will often seems to get in the way of freedom of choice. Many of us have had the experience of being presented with an option so tempting that it felt impossible to resist, and this despite having resolved to do otherwise, or despite our better judgment. Of course, temptation is rarely so strong that it simply cannot be resisted. Nonetheless, in many circumstances, temptation makes it much harder to choose certain courses of actions over others. And in such circumstances, there may be uncertainty as to whether we will succeed in carrying our intentions through, if we resolve to do so. The question then arises, how does one factor in weakness of will when assessing how much freedom of choice agents are afforded by given sets of alternatives?

Standard approaches to the measurement of freedom are surprisingly poorly equipped to answer this question. Numerous philosophers

* For their insightful comments, my thanks to two anonymous reviewers, Nicholas Baigent, Victoria Barham, and the audience of the London School of Economics Choice Group. Very special thanks to Alex Voorhoeve, Campbell Brown, and Richard Holton, whose frequent feedback and advice were invaluable to me.

Ethics 130 (April 2020): 384–414
 0014-1704/2020/13003-0004$10.00

have of course argued, persuasively, that motivational foibles like weakness of will are fetters on human freedom, most notably Charles Taylor.[1] However, the implications of this insight have largely gone unnoticed among those theorists who are interested in the measurement of freedom. In particular, although much has been written on how preference orderings over given sets of options and the degree of dissimilarity between these options affect the amount of freedom of choice that they offer, the thought that different options that are available might be more or less accessible has received little attention in the measurement literature. Indeed, as I will show below, existing approaches by and large simply rule out the possibility of unequally accessible options. As a result, these approaches overestimate how much freedom of choice certain sets of alternatives offer. My concern in this article will be to outline a new approach to the measurement of freedom that takes seriously differential option accessibility due to weakness of will.

I proceed as follows. In Section II, I review some of the more influential approaches to the measurement of freedom of choice. These approaches share a common feature: the way in which sets of alternatives are defined implies that they admit of no degrees in accessibility other than fully accessible or nonaccessible, and this will imply that two agents with the same nominal set of options are equally free. In Section III, I turn to the problem of weakness of will. I explore a series of examples in which the strength of an agent's will makes a crucial difference to how free they are. This turns out to be a sore point for existing approaches to the measurement of freedom because they are unable to take account of the impact of strength of will on freedom. In Section IV, I discuss and rebut three objections to my claim that weakness of will reduces freedom by making options less accessible. The first alleges that "internal" constraints like weakness of will are not relevant to the measurement of freedom. The second alleges that the effect of weakness of will is to make our options worse, not less accessible. The third alleges that the effect of weakness of will is to make certain courses of action unavailable to us, not merely less accessible. In Section V, I propose a new way of representing the set of alternatives that is available to agents, which incorporates a measure of the degree of accessibility of each option. Finally, in Section VI, I show how my framework may be used in real-world cases by applying it to issues regarding the ethics of nudge.

II. STANDARD APPROACHES TO MEASURING FREEDOM

Most discussions of freedom of choice begin with MacCallum's analysis of the concept of freedom, according to which when we say that someone is

1. Charles Taylor, "What's Wrong with Negative Liberty," in *Philosophy and the Human Sciences: Philosophical Papers* (Cambridge: Cambridge University Press, 1985), 2:211–29.

free in some respect, we have in mind the following triadic relation: *x* is free from *y* to do/be/have *z*.[2] The *z* parameter denotes agent *x*'s *opportunity set*, that is, the set of all mutually exclusive options that are open to *x* in the present context. The *y* parameter ranges over all the possible constraints that can apply to *x* and prevent *x* from doing, having, or becoming something.

Different conceptions of freedom are supplied by different interpretations of the parameters. Whether we interpret *x* as denoting an individual or a group agent will give us a conception of individual or group freedom, for instance. And different conceptions of individual freedom turn primarily on which constraints are believed to be relevant. For example, the pure negative conception of freedom, as defended by Carter, Miller, Oppenheim, Steiner, Kramer,[3] and others, holds that one is free to do *x* just in case no one actually imposes constraints on us that make *x*-ing impossible. Capability theorists like Sen, meanwhile, consider that one is free to *x* just in case one can, in fact, *x*—that is, just in case there are no constraints at all that would prevent one from *x*-ing.[4] I happen to agree with Sen (for reasons that will become clearer in Sec. IV.*A*), but for our purposes it doesn't much matter which side one takes on this issue because the measurement approach I develop can be adapted to either view.

Indeed, insofar as we are interested in measuring how much freedom of choice individuals have, it is the *z* parameter that is salient, and the sorts of questions that matter are as follows: Given what options are available to you (i.e., not ruled out to you by whatever constraints we think are relevant), and given what options are available to me, which of the two of us has more freedom of choice? And, moreover, given that you are freer than I, how much more freedom do you have? What we're interested in, then, is developing a ranking rule that will tell us, for any two opportunity sets, which one offers the most freedom.

The standard approach to defining such a rule begins by defining *X*, the (finite) set of all options, assumed to be all mutually exclusive. I underscore the importance of this assumption: two options are mutually exclusive if they cannot be performed in conjunction, if the individual cannot choose to do both. When we say that individuals lack free speech

2. G. C. MacCallum, "Negative and Positive Freedom," *Philosophical Review* 76 (1967): 312–24.

3. Ian Carter, *A Measure of Freedom* (Oxford: Oxford University Press, 1999); David Miller, "Constraints on Freedom," *Ethics* 94 (1983): 66–86; Felix Oppenheim, *Political Concepts: A Reconstruction* (Oxford: Blackwell, 1981); Hillel Steiner, "How Free: Computing Personal Liberty," *Royal Institute of Philosophy Lectures* 15 (1983): 73–89; Hillel Steiner, *An Essay on Rights* (Oxford: Blackwell, 1994); Matthew Kramer, *The Quality of Freedom* (Oxford: Oxford University Press, 2003).

4. A. K. Sen, "Welfare, Freedom, and Social Choice: A Reply," *Recherches Économiques de Louvain* 56 (1990): 451–85.

under totalitarian regimes, we do not mean to imply that there are specific utterances that they are prevented from making, but rather that they cannot, say, criticize the government and then also go on to pursue their life as normal. What makes them less free than citizens of liberal democracy is that certain conjunctively possible courses of action that are possible for free citizens are not possible for them. Likewise, if you are being mugged at gunpoint and have to choose whether to give up your money or your life, the one thing you cannot do is walk away with your life and your money. Accordingly, numerous theorists have stressed that options must be understood as conjunctively possible courses of actions that an individual can complete in sequence,[5] and of course any two different sequences of conjunctively possible actions define two different options.

An opportunity set, being any set of (mutually exclusive) options that might jointly be available to someone, is then defined as any nonempty subset of X. We denote by Z the set of all such subsets. We can now define an ordering relation over Z which tells us, for any two opportunity sets, which set offers more freedom of choice. More precisely, we define $\succsim$, a transitive, nonsymmetric, and reflexive binary relation on Z, such that for any two opportunity sets A and B, $A \succsim B$ is interpreted as "A offers at least as much freedom as B," and the aim now is to impose intuitively plausible restrictions on $\succsim$ and thus deliver an ordering over Z.

This approach has been most thoroughly developed by Pattanaik and Xu.[6] The different ranking rules they propose illustrate the many ways one can use this framework to measure freedom of choice. In their original model, for example, they propose three restrictions on the "offers at least as much freedom as" relation.[7] First, anyone who only has one option has no freedom of choice. Second, one always has more freedom of choice with two options than just one option. This is a very weak version of the fairly natural intuition that more options means more freedom of choice. Third, giving two people an identical additional option does not alter how much freedom they have relative to one another: if John is freer than Smith to begin with, then John remains freer than Smith

5. See, e.g., Carter, *Measure of Freedom*, chap. 7; Prasanta K. Pattanaik and Yongsheng Xu, "Freedom and Its Value," in *The Oxford Handbook of Value Theory*, ed. I. Hirose and J. Olson (Oxford: Oxford University Press, 2015), 356–80.

6. See, e.g., Prasanta K. Pattanaik and Yongsheng Xu, "On Ranking Opportunity Sets in Terms of Freedom of Choice," *Recherches Économiques de Louvain* 56 (1990): 383–90; Prasanta K. Pattanaik and Yongsheng Xu, "On Preference and Freedom," *Theory and Decision* 44 (1998): 173–98; Prasanta K. Pattanaik and Yongsheng Xu, "On Diversity and Freedom of Choice," *Mathematical Social Sciences* 40 (2000): 123–30; Prasanta K. Pattanaik and Yongsheng Xu, "Ordinal Distance, Dominance, and the Measurement of Diversity," in *Rational Choice and Social Welfare: Theory and Applications: Essays in Honor of Kotaro Suzumura*, ed. Prasanta K. Pattanaik et al. (Berlin: Springer, 2008), 259–69; Pattanaik and Xu, "Freedom and Its Value."

7. Pattanaik and Xu, "On Ranking Opportunity Sets."

if we give them both the option (which they previously lacked) to go skydiving. Each of these three assumptions seems initially plausible, but the only ranking rule over Z which satisfies all three is the "simple cardinality rule": for any two opportunity sets A and B, $A \gtrsim B$ if and only if A contains at least as many elements as B.

As Pattanaik and Xu themselves point out, this rule is naive—indeed, they present it as a kind of impossibility result. Nevertheless, this rule provides a very natural starting point for measures of freedom, insofar as having more freedom of choice seems to be a matter of having more choices (of a suitable kind, perhaps), and indeed this rule has been the prototype for every rule that has followed. Crucially for my purposes, it has a key feature in common with every other model that has been proposed since: for any possible option x and any opportunity set A, either x is a member of A, or x is not a member of A—opportunity sets are defined as classic sets, and set membership is bivalent. Of course, the bivalence of set membership implies that if you and I can choose from the exact same options, then we are equally free; this, I will argue, is problematic.

Newer approaches have, of course, been proposed which improve significantly on the simple cardinality rule. For example, Pattanaik and Xu note that the simple cardinality rule seems objectionably insensitive to the quality of the options one has to choose from.[8] Here they explicitly follow Sen, who suggests that if you already have interesting life options that are open to you, your freedom of choice does not increase if you are given the additional option of being beheaded at dawn.[9] Likewise, if you are at a car dealership and are hesitating between several options, Sen claims that your freedom does not expand if you are now given the option of buying a car that is identical to one already on offer, except for a defective gearbox. The lesson Sen draws from this is that an individual's preferences over the options that are available to them matter to how much freedom those options offer them. In consequence, numerous authors have proposed to impose some sort of minimal desirability criterion on opportunity sets. Pattanaik and Xu,[10] for instance, propose that an additional option should only be counted as freedom expanding if and only if it is such that at least some rational persons could prefer it to its available alternatives.[11] This restriction looks like a clear improvement over the

8. Pattanaik and Xu, "On Preference and Freedom."

9. A. K. Sen, "Markets and Freedoms: Achievements and Limitations of the Market Mechanism in Promoting Individual Freedoms," *Oxford Economic Papers* 45 (1993): 519–41.

10. Pattanaik and Xu, "On Preference and Freedom," 182.

11. For further discussion, see also Peter Jones and Robert Sugden, "Evaluating Choice," *International Review of Law and Economics* 2 (1982): 47–65; Clemens Puppe, "An Axiomatic Approach to 'Preference for Freedom of Choice.'" *Journal of Economic Theory* 68 (1996): 174–99; James E. Foster, "Freedom, Opportunity, and Well-Being," *Handbook of Social Choice and Welfare* 2 (2011): 687–728; Phillipe van Parijs, *Real Freedom for All: What (If Anything) Can Justify Capitalism?* (Oxford: Oxford University Press, 2005), chap. 1, n. 39.

previous rule, but it does not change the fundamentally bivalent character of opportunity set membership.

This feature remains even when ranking principles are refined in other directions. For instance, numerous authors have emphasized that one's degree of freedom depends crucially on how diverse one's options are. This point has been emphasized by Charles Taylor,[12] who argues that freedom only matters to us because we are purposive beings, and so one should only count as freer than another if one is free to pursue a greater range of purposes, and if one is free to pursue more important purposes.[13] Accordingly, various authors have proposed ways of ranking sets according to how diverse they are. The simplest such proposal comes from Pattanaik and Xu,[14] who essentially propose to rank opportunity sets by how many types of options they contain, where an option type is a set of available options that are all similar to one another.[15] This approach again seems to improve on the simple cardinality rule, but bivalence remains an implicit assumption.

The key problem with the general approach in this literature was anticipated by Isaiah Berlin, when he claimed that the extent of an individual's freedom depends not only on "how many possibilities are open to them" but also on "how easy or difficult each of these possibilities is to actualize," and "how far they are closed and open by deliberate human acts."[16] One way of cashing out Berlin's insight is to say that options may be more or less accessible to individuals, even as they are available to them. The problem that this lesson poses for the traditional models comes out clearly when we consider cases of weakness of will.

III. THE PROBLEM POSED BY WEAKNESS OF WILL

A. What Is Weakness of Will?

We all frequently confront temptation. When we succumb to it contrary to our better judgment, according to the orthodox view on the subject, we display weakness of will.[17] More precisely, according to Davidson's

12. Taylor, "What's Wrong with Negative Liberty."

13. See also Richard Arneson, "Freedom and Desire," *Canadian Journal of Philosophy* 15 (1985): 425–48; Joseph Raz, *The Morality of Freedom* (Oxford: Oxford University Press, 1988), 408–9.

14. Pattanaik and Xu, "On Diversity and Freedom of Choice."

15. For further interpretations, see also Pattanaik and Xu, "Ordinal Distance"; Sebastien Bervoets and Nicolas Gravel, "Appraising Diversity with an Ordinal Notion of Similarity: An Axiomatic Approach," *Mathematical Social Sciences* 53 (2007): 259–73; Martin Van Hees, "Freedom of Choice and Diversity of Options: Some Difficulties," *Social Choice and Welfare* 22 (2007): 253–66.

16. Isaiah Berlin, *Four Essays on Liberty* (New York: Oxford University Press, 1969), 130.

17. See, e.g., Michael Smith, "Rational Capacities; or, How to Distinguish Recklessness, Weakness, and Compulsion," in *Weakness of Will and Practical Irrationality*, ed. S. Stroud

seminal analysis, weakness of will is an intentional action, motivated by some strong present urge, yet contrary to the individual's all-things-considered better judgment.[18] A standard case of weakness of will, on this account, is that of a dieter, who judges that she ought to stick to her diet but nonetheless splurges on cream puffs and cheesecake. Another is a smoker who judges that he ought to quit but still reaches for the next cigarette.

Discussions on the impact of weakness of will on individual freedom generally presuppose this orthodox view of weakness of will. However, as I will argue later, this makes the resulting analyses undesirably narrow, as there are other plausible views on the table. Richard Holton, for example, has articulated a heterodox but nonetheless influential view according to which weakness of will really consists in overreadily or irrationally reconsidering one's resolutions,[19] where a resolution is understood as a future-directed intention which is formed precisely for the purpose of overcoming one's anticipated future inclinations to act contrary to one's intention. On this view, it is irrational to revise one's resolutions if we do so under the pressure of just the urges which our resolution was formed to defeat, and no new decision-relevant information has come to light.

Holton's view is partly motivated by the fact that many of what we consider to be paradigmatic cases of weakness of will are typically preceded by what Holton calls "judgment shift": those who give in to their urges and break their resolutions typically first come to judge that it is better, after all, to do so.[20] A common example: you swear to quit smoking, hold fast for a while, and then an oblivious coworker offers you a smoke; at first you resist, but then you say to yourself, "Ah, well, why not just one? One cigarette won't kill me. In fact, I'd enjoy it, and I can still quit tomorrow!" Then all resistance fades, and you cave to the urge to smoke. This smoker clearly displays weakness of will, and yet he does not appear to be acting against his better judgment.

To avoid committing myself in advance to any particular view, I will consider "weakness of will" to be a cluster concept, covering a range of irrational motivational foibles. All I will assume is what empirical psychology tells us: first, that weakness of will (or "failure of self-control," among

---

and C. Tappolet (Oxford: Oxford University Press, 2003), 18–28; Philip Pettit, "Akrasia, Collective and Individual," in Stroud and Tappolet, *Weakness of Will*, 68–96; Sarah Stroud, "Weakness of Will and Practical Judgment," in Stroud and Tappolet, *Weakness of Will*, 121–46.

18. Donald Davidson, "How Is Weakness of the Will Possible?," in *Moral Concepts*, ed. Joel Feinberg (Oxford: Oxford University Press, 1969), 21–42.

19. Richard Holton, *Willing, Wanting, Waiting* (Oxford: Oxford University Press, 2006), 77.

20. Ibid., 97.

psychologists[21]) manifests as a conflict between some present urge and some antecedent motivation (better judgment, resolutions, etc.) and may hinder our ability to act on certain intentions, and second, that it is typically cue driven, in that these urges usually do not form spontaneously, but rather in response to environmental stimuli.[22] Addictive desires are the most vivid case in point: addicts become sensitized to certain cues associated with past drug use, and subsequent exposure to those cues immediately results in very strong motivational responses, even in former addicts who have not used for a long time.[23] This is why addicts who have long been sober are much more likely to use drugs in the future than individuals who have never used drugs before. Also, more salient cues typically trigger stronger urges.[24]

By contrast, strength of will—the capacity to recruit the motivation to defeat one's present urges—is generally thought to be enabled by the use of a distinctive cognitive capacity, willpower, which is effortful to employ but enables us to enforce our better judgment and commitments in the face of contrary desires.[25]

There are two dominant models in the experimental psychology literature on how this faculty operates. The "cognitive depletion" model of willpower, pioneered by Baumeister and Heatherton,[26] interprets willpower as a kind of cognitive resource which we call upon at will to recruit the motivation to carry through our resolutions. This resource is limited in supply over the short term and "spent" in resisting temptation (this is felt as the exertion of effort), so that previous use of our willpower leaves less available for subsequent use in the short term. Individuals can thus progressively exhaust their willpower, making future attempts at resoluteness less likely to succeed. The phenomenon of "ego depletion" is often cited as providing evidence for this view. To cite just one example, the experiments of Baumeister and Heatherton seemed to show that forcing oneself to resist temptation in one period makes one less likely to persist in one's efforts to complete a difficult or frustrating but completely unrelated task in the next period. In their initial study, forcing oneself to eat radishes instead of the more tempting chocolates made one less likely, in the subsequent period of the experiment, to persist

21. See, e.g., Drew Fudenberg and David K. Levine, "Timing and Self-Control," *Econometrica* 80 (2012): 1–42.

22. Richard Holton and Kent C. Berridge, "Addiction between Compulsion and Choice," in *Addiction and Self-Control*, ed. Neil Levy (Oxford: Oxford University Press, 2013), 239–68.

23. See Terry E. Robinson and Kent C. Berridge, "Addiction," *Annual Review of Psychology* 54 (2003): 25–53.

24. Holton, *Willing, Wanting, Waiting*, 97.

25. Ibid., 112.

26. R. F. Baumeister and T. Heatherton, "Self-Regulation Failure: An Overview," *Psychological Inquiry* 7 (1996): 1–15.

in one's efforts to solve puzzles. This was taken to show that willpower is depleted with use.

This is not to say that every case of abandoning one's resolutions is a case of having exhausted one's willpower; it is surely quite common for people to give in to temptation without ever calling upon their willpower to resist the urge to yield ("I know I resolved not to eat chocolate cake, but things are different today: I'm on vacation!"). As Rizzo puts it, willpower is just another scarce resource: we can choose how to allocate it across different consumption bundles, but there is a hard cap on how much we can buy.[27] Importantly, individuals can become better at exercising their willpower. According to Baumeister and Vohs, the more one succeeds in being resolute, and the more one cultivates certain habits of plan making, the easier it becomes to recruit motivation going forward (if one is keeping the resource metaphor in mind, training can give you more motivation-recruiting bang for your effort-of-will buck).[28] This interpretation has many adherents in behavioral economics and was until recently the orthodox view.[29]

In recent years, however, the "attentional myopia" model of willpower has emerged as a major rival theory. On this interpretation, weakness of will arises in situations when our attention narrows on the most salient cues (due to cognitive stress or prior sensitization), which suggest resolution-violating behaviors.[30] Cognitive load makes us susceptible to temptation, on this view, not because complex cognition consumes willpower but because the way that we typically deal with cognitively demanding situations is by focusing our attention on the most salient features of that situation.[31] So, for example, one experiment conducted by Mann and Ward found that dieters who performed cognitively demanding tasks consumed twice as much milkshake right afterward when placed in a room containing salient food items than when in a room containing conspicuously placed scales and dieting books, whereas dieters who performed cognitively low-load tasks were significantly less influenced by cues.[32]

Strength of will is interpreted on this view not as a cognitive resource but as a skill in refocusing one's attention. Elkins-Brown, Teper, and Inzlicht

27. Mario J. Rizzo, "Behavioral Economics and Deficient Willpower: Searching for Akrasia," February 12, 2016, https://ssrn.com/abstract=2731818.

28. Roy F. Baumeister and Kathleen D. Vohs, eds., *Handbook of Self-Regulation: Research, Theory, and Applications* (New York: Guilford, 2004).

29. See Kathleen D. Vohs and Todd F. Heatherton, "Self-Regulatory Failure: A Resource-Depletion Approach," *Psychological Science* 11 (2000): 249–59; Rizzo, "Behavioral Economics"; Fudenberg and Levine, "Timing and Self-Control."

30. See, e.g., Michael Inzlicht and Brandon Schmeichel, "What Is Ego Depletion? Towards a Mechanistic Revision of the Resource Model of Self-Control," *Perspectives on Psychological Science* 7 (2012): 450–63; Traci Mann and Andrew Ward, "Attention, Self-Control, and Behaviour," *Current Directions in Psychological Science* 16 (2007): 280–83.

31. Mann and Ward, "Attention, Self-Control, and Behaviour," 281.

32. Ibid., 281, 282.

argue that it is possible to recruit the additional motivation to maintain one's resolutions in the face of temptation by refocusing one's attention on cues which prompt self-control (e.g., to think about how the action would be socially inappropriate), in particular through exercises of mindfulness, which make one more receptive to cues which would otherwise lack salience.[33] It has been found, for example, that those who practice meditation experience better outcomes in smoking cessation and alcohol use than control groups,[34] and there is some evidence to suggest that meditation may be an effective treatment for drug abuse disorders.[35] Elkins-Brown, Teper, and Inzlicht argue that this is because "mindfulness meditators are better prepared to acknowledge moment-to-moment affect that signals the need for self-control."[36]

There is substantial evidence for both dominant views, but regardless of which model has got the details right, there are some broad-stroke conclusions that come from the experimental psychology literature. First, willpower is a cognitive resource or skill which we can call upon to recruit the motivation necessary to defeat resolution-inconsistent urges. Second, some agents are better than others at leveraging this faculty. This is well attested to by the fact that individuals can train themselves to better resist temptation, for example, by practicing mindfulness exercises, or cultivating habits of plan making. Third, cognitively demanding tasks and exposure to temptation temporarily undermine one's ability to subsequently recruit additional urge-defeating motivation.[37]

B. Problem Cases

Let us now look at some cases. Consider first Weak-Willed Willy and Resolute Regina. Suppose that every night for a whole week both have the choice between four alternatives: completing job applications, solving

33. N. Elkins-Brown, R. Teper, and M. Inzlicht, "How Mindfulness Enhances Self-Control," in *Mindfulness in Social Psychology*, ed. J. C. Karremans and E. K. Papies (New York: Psychology, 2017), 65–78.

34. See H. M. Elwafi et al., "Mindfulness Training for Smoking Cessation: Moderation of the Relationship between Craving and Cigarette Use," *Drug and Alcohol Dependence* 130 (2013): 222–29; Y. Y. Tang, R. Tang, and M. I. Posner, "Brief Meditation Training Induces Smoking Reduction," *Proceedings of the National Academy Sciences* 110 (2013): 13971–75; S. Bowen et al., "Mindfulness Meditation and Substance Use in an Incarcerated Population," *Psychology of Addictive Behaviors* 20 (2013): 343–47.

35. A. Chiesa and A. Serretti, "Are Mindfulness-Based Interventions Effective for Substance Use Disorders? A Systematic Review of the Evidence," *Substance Use and Misuse*, 49 (2014): 492–512.

36. Elkins-Brown, Teper, and Inzlicht, "How Mindfulness Enhances Self-Control," 67.

37. See M. Muraven, D. M. Tice, and R. F. Baumeister, "Self-Control as Limited Resource: Regulatory Depletion Patterns," *Journal of Personality and Social Psychology* 74 (1998): 774–89; Mann and Ward, "Attention, Self-Control, and Behaviour"; Holton, *Willing, Wanting, Waiting*, chap. 4.

logic problem sets, working on their essays, and going out to drink with friends. In other words, their opportunity sets are identical. Both would love nothing more than to go out with their friends, but they have resolved not to because this would adversely affect their capacity to meet urgent deadlines with respect to their work, and they both judge that meeting these deadlines is more important than enjoying themselves that week. Regina, being resolute, is pretty consistently successful in recruiting the motivation to resist temptation, and she is very skilled at shifting her attention toward cues that facilitate self-control. She is quite reliable: if she says she will do something, or that she judges that she ought to do it, it is generally a sure bet that she will. Willy, by contrast, is a flake, generally unsuccessful in resisting temptation when it presents itself, and unskilled at refocusing his attention. He is almost always carried away by his present urges to renege on his previously stated intentions and judgments.

The first thing to observe here is that on standard constructions of opportunity sets, Willy and Regina are equally free. In every relevant respect, after all, their opportunity sets are the same: the options in both sets are equally many and diverse, they are all desirable, and they present Willy and Regina with the exact same sorts of opportunities. There's nothing one can do that the other cannot (it may be more or less difficult, or more or less pleasant, but they both can do it). Thus, on the traditional picture, they simply cannot be unequally free. But this looks like a mistake. Indeed, whatever the outcome of this case, I submit that Willy is less free than Regina: either he expends his willpower reserves more quickly, less efficiently than she does in attempting to be resolute, or he is simply far less skilled at shifting his attention to recruit the motivation to defeat his urges. Sure, he may well succeed in standing firm in the face of temptation, but there is a clear sense in which he is much less capable of doing so than she is.

But in what sense does this make Willy less free? Charles Taylor argues that we "experience our desires and purposes as qualitatively discriminated, as higher or lower, noble or base, integrated or fragmented, significant or trivial, good and bad."[38] And what it is for an end to be qualitatively superior (inferior), on his view, is for it to be the sort of end that we desire to (not) desire. Of course, as he observes, what we actually desire and what we desire to desire can come apart; thus, we may on occasion pursue base goals instead of ones we deem worthier—this is just what it is to be weak-willed. But to be driven by one's defects to pursue base ends is ipso facto to be constrained in one's ability to pursue the ends one truly finds worthwhile, and this is injurious to our freedom, Taylor insists.[39]

38. Taylor, "What's Wrong with Negative Liberty," 220.

39. There are other influential views one could consider on how internal constraints impair individual freedom. Crocker and Christman, for instance, both argue that heteronomy

Taylor seems to me to have gotten something right here: motivational defects like weakness of will constrain freedom because they make us less capable of pursuing certain aims. That said, this on its own remains quite vague (what is it for a person to be more or less capable, for instance?), and in some respects Taylor's analysis lacks generality. For one thing, it is very much tied to the orthodox view of weakness of will and cannot be extended to nonstandard views. If, as Holton argues, weakness of will is often characterized by the reevaluation of what we are initially disposed to regard as base desires into acceptable and even worthy ones, then it cannot be in those cases that our freedom is constrained because we are unable to pursue ends we find worthwhile. Rather, in such cases it is our ability to carry out our resolutions that is impaired.

This point bleeds into a further one. For Taylor, weakness of will is a restriction on freedom only because it makes us choose lesser options over better ones; this analysis might seem too value laden. It suggests a view on which something only constrains our freedom if it make us worse off, in some way. Isn't the problem more generally just that susceptibility to temptation undermines one's ability to act on certain aims and intentions, regardless of whether those are ones we strongly identify with? A workaholic might desire to desire to work; nonetheless, workaholic compulsions plausibly impair one's freedom to relax and go on vacation. A more value-neutral analysis of constraints is needed to account for these kinds of cases, one which allows constraints to be either good or bad and only requires that they make one less capable (we will return to the issue of value neutrality in Sec. IV.*B*).

In light of the above considerations, I propose the following analysis of Willy's and Regina's situation. Willy is less free than Regina because although all four options are available to both of them, Willy's work options are less accessible to him, in the sense that, conditional on his intending to work, he is less likely to successfully enforce his intention and actually work than Regina is, conditional on her intending to work. It is as though, in forming the intention to work, Willy and Regina both take out a lottery ticket which gives them a particular chance of successfully working and a particular chance of failing and choosing to go out instead; because Willy is weak-willed, his ticket gives him longer odds of

---

impairs freedom, and that the less autonomous one's decision to φ, the less free one was not to φ. Taylor's analysis is particularly salient for us, however, because his account is explicitly concerned with explaining how weakness of will constrains individual freedom, unlike Crocker's or Christman's. And indeed, their analyses have no obvious application to the present case. The paradigm cases of heteronomy Crocker considers are cases of impaired consciousness and coercion, while for Christman the autonomous character of a decision depends on the desires driving the decision being formed by a process the individual would not have resisted had they attended to it; by either author's lights, Willy and Regina might both count as fully autonomous. Lawrence Crocker, *Positive Liberty* (London: Nijhoff, 1980); John Christman, "Liberalism and Individual Positive Freedom," *Ethics* 101 (1991): 343–59.

success. This analysis is more precise than Taylor's, as it cashes out in terms of conditional probabilities the sense in which Willy is less capable than Regina of enforcing practical attitudes. But it is also more general, as it is plainly value neutral,[40] and it does not assume a particular picture of weakness of will, only that it is a foible which may hinder our ability to act on certain intentions, but which may be overcome by an effort of willpower, and which some are more adept at overcoming.

This analysis also happens to be congruent with the sort of analysis which David Manley and Ryan Wasserman argue ought to be made of all dispositional properties.[41] A dispositional property is the property of behaving in a certain way if certain conditions obtain. More precisely, Manley and Wasserman argue that an object is disposed to behave in a certain way when a given stimulus condition obtains if and only if the object would behave in that way in suitably many cases where the stimulus condition obtains.[42] And accordingly, one object is more disposed to behave in a certain way than another if and only if the first object would exhibit the behavior of interest in more cases where the relevant stimulus condition obtains than the second object would. Importantly, some dispositions require no specific stimulus conditions: if one is prone to anger in any situation whatever, one is irascible. Being strong-willed is arguably a similar sort of disposition: it is to be reliably disposed to act on one's intentions

40. Carter (*Measure of Freedom*, 153–55) discusses a proposal very similar to mine, on which we must assign numerals between 0 and 1 to each option we are externally unconstrained from performing, based on the degree of internal constraint, with "1" representing complete internal freedom to perform it and "0" representing the complete internal unfreedom to perform it. Carter dismisses this proposal on the grounds that the only way to distinguish internally constraining and nonconstraining desires is by reference to the value of the purposes that these desires hinder. Since I defend a version of this view, I feel I must make a couple of points in response. First, it is unclear in Carter's treatment how the numbers are to be interpreted. He is led to consider this view by the thought that we ought to measure "the internal unfreedom of an agent to do x in terms of her propensity to do some other thing," which invites a probabilistic interpretation of the numbers. However, he goes on to argue that the degrees of internal constraint represented by the numbers "would appear to correspond to the 'strength of will' that is required to overcome them," and he quotes Flathman's taxonomy of involuntary behaviors, ordered by their degree of involuntariness, as offering a possible measure of the degree of internal constraint on an individual. This suggestion is nonprobabilistic and leaves mysterious how one arrives at a cardinal measure of the strength of efforts of will. My approach has the virtue of clarity: internal constraints have an unambiguous interpretation in terms of conditional probabilities. Second, and more importantly, it is false that a notion of internal constraint must be objectionably value laden: my own proposal makes no reference to values or preferences, only intentions and conditional probabilities. Carter is therefore too quick to reject this approach. See Richard Flathman, *The Philosophy and Politics of Freedom* (Chicago: University of Chicago Press, 1987).

41. David Manley and Ryan Wasserman, "On Linking Dispositions and Conditionals," *Mind* (2008): 117: 59–84.

42. Ibid., 76.

(whatever those are) in suitably many cases, notably in cases of temptation. It would follow, on Manley and Wasserman's analysis, that to be more or less strong-willed is just to be more or less likely to successfully enforce one's intentions, whatever those are. This is just what I have argued. Thus, a general analysis of dispositions supports my view of why weakness of will is a constraint on freedom: it depresses one's chances of enforcing one's intentions.

Note that we can extend my analysis to other cases. Consider a person suffering from intense depression who is trying to hide her condition. Such a person can, usually, muster the will to go to work, attend social events, go out with her friends, and so on; it can feel like drowning, but she may succeed in hiding her condition if she is determined. Sometimes, however, her will fails, and she will wake up to find that she cannot get out of bed. She may believe that the best thing for her right now is to try to maintain a "normal" social life, and she may know that staying in bed all day will only make her feel worse about herself, but still she fails to recruit the motivation to unmoor herself from her bedposts. Our patient is arguably not weak-willed—she is not driven by some strong desire to act against her better judgment or to break her resolutions—but rather prey to what Aquinas called accidie, or the total loss of motivation. Nonetheless, she would rightly view the end of her depression as a liberation, and for the same reason as before: not because it affords her any more options, but because it restores to her control over her choices. Once it is no longer a battle of uncertain outcome to recruit the motivation to get out of bed, to go to work, or to make dinner, she can once again count on herself to act on her practical attitudes (intentions, resolutions, better judgments, etc.).

Going further, my analysis suggests that Willy might have more freedom if his friends were out of town, leaving him with only the three work options. Free from temptation, free from the constant, painful battle to maintain his resolution which he was not guaranteed to successfully carry out, he now has no trouble at all setting priorities between work tasks and delivering quality work on time for all his deadlines. The loss of his tempting option may therefore leave Willy with more freedom, by making his remaining options more accessible. This is difficult to make sense of under standard approaches to the measurement of freedom: there is nothing that Willy can do now that he could not do before, no outcome that he can achieve now that he could not achieve before—much to the contrary: there is less he can do now—so it looks very much as though we have simply removed an option from him, and standard monotonicity assumptions imply that this never increases one's freedom.[43]

43. Pattanaik and Xu ("On Preference and Freedom") are an exception here; if an option dominates every other option in a set under every rational preference ordering, then

We can draw some preliminary lessons here. The degree of freedom offered by one's opportunity set depends, first, on whether choosing any of the options on offer requires one to recruit any motivation to defeat one's urges to choose otherwise; second, on how strong-willed the agent is, that is, how capable they are of recruiting motivation; and third, on what other options are available. A fourth important factor is the salience of each option. This is brought out clearly in the case of addiction.

Consider Perseus and Cassandra, two cocaine addicts who successfully complete their detox program and return to their homes. They both resolve never to do drugs again, they are equally strong-willed, and they face the same constraints, so that the same options are available to them. Anything the one can do, the other can as well. We can even assume that they share the same preferences and value the same alternatives. On any standard measure of freedom, they are equally free. The only difference between the two is that while Perseus still lives in his old flat where he is regularly exposed to drug cues, Cassandra has the foresight to move to a part of town where she does not face any such cues (though it is a short drive to a dealer). In other words, Cassandra and Perseus have the same (or very nearly the same) opportunity set, yet I contend that Cassandra will have more freedom of choice when she gets home from detox than Perseus.

Indeed, given that Perseus is exposed to various drug cues, the possibility of doing cocaine will be very salient to him, triggering his addictive desires. Plainly, the attentional myopia model of willpower predicts that his attention will narrow on the salient behavioral cues in his environment, which are very suggestive of cocaine, especially if he is ever under conditions of cognitive stress. And on resource models of willpower, resisting the pull of his addictive desires will very seriously deplete his willpower reserves, requiring a great deal of effort. On both models then, resisting his addictive desires is extremely difficult, and he is not all that likely to succeed anyway. In contrast, since Cassandra is not exposed to these cues, her addictive desires are not triggered, and so she will have no difficulty in maintaining her resolution to stay clean. Hence, despite the fact that Perseus and Cassandra have identical opportunity sets and are equally strong-willed, the simple fact that the "wrong" option is made salient to Perseus and not to Cassandra means that the former has less freedom of choice, and for the same reason as before: because of his addictive desires, he is less capable of making certain choices, and so some of his options are less than fully accessible to him.

---

removing that option might increase one's freedom. Clearly, though, this is not the situation here, so their monotonicity assumptions would likewise imply that removing Willy's fourth option does not increase his freedom.

The point that these examples have labored to bring out is simply this: options admit of degrees in accessibility, and what it is for an option to be more or less accessible is simply for it to be the case that, conditional upon your intending to perform it, you are more or less likely to perform it, successfully enforcing your intention. Motivational defects like weakness of will clearly inject great heterogeneity in people's capability to make choices, and standard approaches to the measurement of freedom are oblivious to these distinctions. To more meaningfully compare individual freedom requires a framework that captures the heterogeneity in the capability of different agents to do, be, and have things worth wanting.

IV. OBJECTIONS

Before going on to propose a new model, I consider three responses to the argument I have just put forward. Some will be skeptical that weakness of will really does impose any constraint on individual freedom. Others might concede that weakness of will is a constraint but argue that this is either because weakness of will somehow impairs the value of our choices or because it actually rules out altogether certain conjunctive courses of action. Let us look at each in turn.

A. Negative Freedom

Recall that proponents of negative conceptions of freedom claim that one is free to φ if and only if one is not prevented from ϕ-ing by another (through deceit, coercion, etc.). There is a very sharp distinction on such views between being unable to do something and being unfree to do it. Nature constrains us in all sorts of ways but does not thereby make us want for freedom. Paraplegics are not unfree to climb mountains, though they cannot do so. Accordingly, a first response to my argument might be that Willy and Perseus are not in any way less free than their female counterparts, nor does depression constrain one's freedom, because weakness of will is a feature of one's psychology, not a barrier to choice imposed on us by others.

As stated, this claim is surely too strongly put. True, in my previous examples, weakness of will seems to act as an internal constraint. But weakness of will can arise through many different processes, some of which have their origin in the deliberate behavior of others, and in those cases it looks much more like an external constraint. Suppose I tempt an addict by waving bags of cocaine under their nose, thereby triggering their addictive desires and making it exceedingly difficult for them to go on about their business as they had planned; is it I or defective brain chemistry that constrains our addict's ability to stay clean? Or suppose I am a terrible boss who creates a hostile, oppressive work environment, leading some of my

employees to become so depressed that they miss days of work; is it I or my employee's own unfortunate but natural tendency to depression that now binds them to their beds?

In cases like these, I act like a polluter: I am introducing temptations and sources of cognitive stress into the social environment, triggering arduously resisted compulsions in others who, but for my behavior, would be considerably more capable to do what they wanted to. This is not so different from how, by poisoning the soil with heavy metals, industrial waste dumping impedes our freedom to build houses, grow crops, raise farm animals, or procure drinking water. These acts of pollution clearly interfere with the decisions of others and constrain what they may do.

In fact, we can quickly construct much more chilling scenarios than ones involving pollution. Imagine, for instance, that a megacorporation uses subliminal messaging or highly aggressive advertising and propaganda to manufacture addiction-like desires in us, which it then triggers at its convenience to manipulate our behavior, in contravention of our better judgments and resolutions. These desires may well be resistible, with enough strength of will, but it is clear that the corporation is interfering with our choices, curtailing our freedom. But what is remarkable in this case is that the corporation is controlling us by exploiting our susceptibility to weakness of will. This and the pollution scenario reveal that even if one insists that the only genuine constraints on freedom are those imposed on us by others, one must concede that there are cases in which weakness of will and related motivational defects like depression and compulsion are constraints imposed on us by others. Thus, at least in those cases, the proponent of negative freedom must care about how weakness of will constrains our freedom, and what my treatment of this question shows is that they must give up bivalence.

Finally, it is not clear to me that the distinction between internal constraints and external constraints is well formed. Suppose I fall into a pit trap and cannot get out. My freedom is curtailed in this case, but is this because I face the external constraint that the pit was dug too deep or because I face the internal constraint that I cannot jump high enough? Likewise, is the constraint Willy faces purely an internal one? He only struggles to be resolute, after all, because he's tempted by his friends' decision to go out to the pub. It looks as though whenever a person is unfree to do something, this is always due to the joint impact of internal and external factors. Thus, external factors, in particular, cannot by themselves prevent anyone from doing anything; they require the cooperation of internal factors. This makes it difficult to articulate a coherent picture on which one is free to φ just in case there are no "external constraints" on one's φ-ing. For this reason, I submit that it is better to side with Sen and accept that you are free to φ just in case you can, in fact, φ.

This is not to deny that there is a morally salient difference between the situation of a person who is barred from attending university by the state because she is a woman and the situation of a person who simply cannot afford tuition. But the difference, I think, is that one is a worse form of unfreedom than the other, not that one is a case of unfreedom and the other a case of something else. Likewise, to take an example from Kramer,[44] it is much worse to be unable to walk more than a hundred meters because one has been assaulted than to be unable to walk more than a hundred meters because of a disability. There is a relevant contrast between these two cases. But, as Sen emphasizes, there is already a morally salient contrast between being capable of doing something and being incapable of doing it.[45] It is this second contrast which is of interest to us when we aim at a measure of freedom: we want to know how capable people are. And with this in mind, the problem posed by weakness of will becomes particularly salient, as weakness of will does constrain individuals' capabilities, however it arises.

B. The Value of Options

Of course, one might accept everything I have just said but contest my explanation of how weakness of will constrains individual freedom. Numerous theorists have defended what Carter refers to as a "value-based" account of freedom, according to which one's degree of freedom is determined, in some robust way, by the value of one's options;[46] accordingly, a critic might object that the real reason weakness of will constrains individual freedom is that it somehow impairs the value of one's opportunity set.

Sen, Arneson, Crocker,[47] and others have argued that, all else being equal, an individual has more freedom of choice than another if her options are better. "Better" can be interpreted either subjectively, as reflecting the individual's own preferences or value judgments, or objectively, as reflecting some impartial assessment of how good the options are. As we saw earlier, Taylor deems you freer than I if you can pursue the aims you desire to desire and I cannot, so his is a subjective interpretation. Sen, similarly, argues that if you and I both only have one option, but you like the one you have and I do not like mine, then you must be seen to enjoy a greater degree of freedom.[48] In contrast, Raz takes a more objective view, arguing that ("positive") freedom is expanded by whatever expands

44. Kramer, *Quality of Freedom*, 367.

45. A. K Sen, *The Idea of Justice* (London: Penguin, 2009), 209.

46. Carter, *Measure of Freedom*, 170.

47. A. K. Sen, "Welfare, Freedom, and Social Choice: A Reply," *Recherches Économiques de Louvain* 56 (1990): 451–85; Arneson, "Freedom and Desire"; Crocker, *Positive Liberty*.

48. Sen, "Welfare, Freedom, and Social Choice."

autonomy, which is expanded in proportion to (among other things) the diversity of options which one has and which allow one to develop all of one's mental and physical abilities.[49]

In any case, if one is going to rely on preferences or value judgments to (partially) determine the degree of freedom offered by a set of options, then it becomes necessary to specify options in terms of all the features that are relevant to our assessment of these options, and here it might be argued that the way in which I specified the options in Section III was incomplete. I alleged that Willy and Regina both have the same work options, but in fact (so the objection goes) the options available to Regina are "work on problem sets without exerting willpower," "work on essay without exerting willpower," and so on, while the options available to Willy are "struggle to overcome temptation and work on problem sets," "struggle to overcome temptation and work on essay," and so on. Accordingly, when Willy's friends leave town, it is not that options which were previously available became more accessible; rather, it is that the options "struggle to *x*," "struggle to *y*," and "struggle to *z*" were replaced with the options "do *x* without struggle," "do *y* without struggle," and "do *z* without struggle."

Redescription on its own does not get us very far, since Willy and Regina still have equally many options to choose from,[50] which seem equally diverse, and it seems rationally permissible for both Will and Regina to choose any of their available options. However, once options are appropriately redescribed in this way, the following explanation for why Willy is less free than Regina becomes available: it is not that his options are less accessible to him than Regina's are to her, but rather more simply that her options are better, insofar as they require less of a struggle. Similar explanations are available for Perseus and Cassandra, as well as for our depressed patient. What addiction and depression do to a person is not to make certain courses of action less accessible to them but merely to make certain courses of action altogether impossible without struggle and torturous effort. Thus, bivalence is saved: either options are available or they are not, full stop, and weakness of will only impacts our freedom by making our options more or less valuable.

49. Raz, *Morality of Freedom*, 409, 376.

50. This may seem surprising at first glance: Willy cannot do *x*, *y*, or *z* without struggling, whereas Regina can, so doesn't it follow that she disposes of more options than Willy? But notice that just as Willy cannot do *x*, *y*, or *z* without struggling, Regina cannot choose to struggle to do *x*, *y*, or *z*—she can only choose to do *x*, *y*, or *z*, without struggling, since there's nothing for her to struggle against—whereas Willy can choose to struggle to do *x*, *y*, or *z*. It is not as though, being strong-willed, she can choose between doing *x* without struggle and doing *x* by struggling. If it requires no effort of you to turn down a friend's offer of a cigarette, you cannot make yourself struggle to turn it down. So both Willy and Regina have a choice of exactly three options, differing only in how strenuous an effort of will they require in order to be carried out.

There are two reasons why it would be a mistake to pursue this strategy. First, tying the degree of an individual's freedom this tightly to the value of their options risks eliding the distinction between the amount of freedom one has and the value of that freedom to us. This is a contrast worth preserving, because in order to have a clear-eyed view of what tradeoffs we are making in weighing freedom against other values, it pays to have a measure of freedom that does not make one's degree of freedom dependent on the extent to which it promotes other values, with which it may compete.[51] The value of one's options of course matters—greatly, in fact—but having more freedom is not primarily a matter of achieving better outcomes; it is a matter of being more capable of doing, having, being, and becoming things which we have (at least minimal) reason to value.

Accordingly, when confronting the case of Willy and Regina, the critical question is not which of the two has the better options, or who achieves the most desirable outcomes, but which of the two is most capable of achieving ends that one might have reason to value. I claim Regina to be more capable. She is the one who has the most control over what she ultimately does: if she judges that deadlines do not really matter, she can choose to go out, but if she judges that she ought to solve problem sets, or resolves to do so, she is capable of reliably enforcing her judgments and commitments. My approach gives us a way of asserting that Regina is freer than Willy without committing ourselves to any judgment regarding the value of their options, and this is an attractive feature of the approach.

Second, even if we do accept the value-based view, it is at most a partial story, because options are not necessarily worse if they are more effortful. Some people may find that there is great value in having to struggle to get what they want, and that having easy choices cheapens the value of what is obtained; they may place greater value on certain options precisely because they require more willpower. Nietzsche, for instance, reserves high praise for those individuals locked in a constant struggle to overcome themselves and their limitations: "That one has become more indifferent to hardship, toil, privation, even to life. The man who has become free . . . spurns the contemptible sort of well-being dreamed of by shopkeepers, Christians, cows, women, Englishmen and other democrats. The

51. Note that the ranking rule mentioned in Sec. II, according to which an option only contributes to your freedom if at least some rational person could prefer it to its available alternatives, while obviously not entirely value neutral, nonetheless preserves this distinction. This view allows an option to count as freedom expanding even if no one actually prefers it, and it explicitly denies that, all else being equal, having better options makes you freer: an individual with three terrific options is exactly as free as an individual with three middling options (or even three bad options), provided that any of the three middling options could be rationally preferred to the other two. Accordingly, this ranking rule would fail to imply that Willy is less free than Regina (see note 3), unless one accepts my proposal to reject bivalence.

free man is a warrior. . . . One would have to seek the highest type of free man where the greatest resistance is constantly being overcome."[52] Willy, likewise, perhaps because he is an avid student of Nietzsche, may spurn the ease with which Regina goes through life, never prey to her passions or experiencing the pain of self-struggle, and he may thus prefer his own more effortful opportunity set. Indeed, Willy may object to any paternalistic intervention that would remove temptation from his sight, arguing that it would rob him of his struggle, of his chance to triumph over himself. Still, I insist that he and individuals like him are, in at least one respect, less free to choose than they would be if they were not required to recruit any willpower in order to choose any of their actions: when Willy reaches out to choose to do what he has resolved to, he is likelier to find his reach too short and choose otherwise in the end. Whatever the impact of an option's effortfulness on that option's desirability, there is in any case a separate impact on how much freedom this leaves one with.

C. Conjunctively Possible Courses of Action

Finally, a critic might object that my analysis misunderstands the mechanics of weakness of will, as they are revealed to us by the experimental literature. More precisely, it might be argued that weakness of will ought not to be understood as a probabilistic constraint on one's ability to carry out certain courses of action conditional on intending to carry them out, but rather as a ruling out of certain classes of conjunctive options.

What the ego depletion experiments purportedly reveal is that if one uses one's willpower in one period to resist temptation, then it becomes much less likely that one will resist temptation in the next period; one possible interpretation of these findings is that one is fundamentally limited in how much motivation one can recruit over a given period of time, with some being more limited than others. So suppose I believe I ought to do x and y, but that I am also very strongly tempted to do z, and that the time frame in which these options are performable is limited in such a way that I can only perform two of x, y, and z in sequence; doing what I think I ought (and have resolved to do) means not doing what I really most want to do. If I have the will to overcome this temptation, then my opportunity set is $\{(x, y), (x, z), (y, z), (y, x), (z, x), (z, y)\}$, but if performing x and y (in any order) requires me to recruit more motivation than I am actually capable of, then in fact my opportunity set is $\{(x, z), (y, z), (z, x), (z, y)\}$.

If we now return to the Willy and Regina scenario, then, on this view, the correct explanation for why Willy is less free than Regina is that since

52. Friedrich Nietzsche, *Twilight of the Idols*, trans. R. J. Hollingdale (Baltimore: Penguin, 1968), TA 92.

Willy is weaker-willed than Regina, there is less motivation that he is able to recruit over any given period of time than she is. His "willpower budget" over a given time horizon is smaller. Therefore, if he recruits the motivation to resist temptation on one night, he will have less willpower left to "spend" going forward than Regina, meaning that some possible courses of action which would be open to her (counterfactually, at least) would not be open to him. In this way, bivalence is saved: weakness does not make options less accessible; it just makes some options unavailable.

In reply to this objection, I would point out that the idea of a literal willpower budget can only be meaningful on the resource model of willpower. On the attentional myopia model, cases like Willy's and Regina's break down in purely probabilistic terms: being more or less skilled at recruiting motivation by refocusing one's attentions just is a matter of being more likely to act on one's resolutions and to follow one's better judgment. If the resource model turns out not to be empirically adequate, then this "budget constraint" idea cannot explain why Willy appears to be less free than Regina. My proposal, however, is compatible with any substantive view of willpower. After all, it is worth reminding ourselves that whether or not they reveal the existence of a literal willpower budget, ego depletion experiments certainly suggest that depleting one's willpower reserves makes it less likely that one will succeed in being resolute going forward.

Perhaps it will be replied here that if one is not guaranteed to succeed in carrying out a particular course of action, then one is not free to carry it out at all, but only free to try to carry it out. On this account, the only difference between Willy and Regina is that one is likelier to carry out courses of action that they are both free to try to carry out. It would follow that Regina really is not freer than Willy to do anything. Weakness of will can only make a difference to how free people are if it makes a difference to how many courses of action they can be guaranteed to carry out if they try.

This would be a very radical reply. Only necessary propositions have a probability 1 of being true, and there is simply no course of action which it is necessarily true that we would succeed in carrying out if we tried. It would follow, on this reply, that none of us are free to do anything, only to try to do things. But anyone can try to do anything: I can try to flap my arms and fly to the seventh moon of Jupiter, and so can you. Are we all equally free then? This would be an unfortunate conclusion. Rather than admit to this, I suggest we simply accept Berlin's view that our freedom depends both on how many possibilities are open to us and on how easy or difficult these possibilities are to actualize—and I propose to explicate the sense in which possibilities can be easier or harder to actualize in terms of conditional probabilities, which requires giving up bivalence.

V. FUZZY FREEDOM

I now present my view on how freedom of choice should be measured. There is actually a simple way of capturing the idea that options may be more or less accessible to an agent, and this is to represent opportunity sets as so-called "fuzzy sets," where the membership of an element in the set is not bivalent. I show below that this will allow us to generalize the models canvassed in Section I so that they yield the intuitively correct rankings in the problem cases laid out in Section II.

But first, let me provide some definitions. I define an opportunity set as a pair (X, m), where X is the set of all options, assumed finite, and m is a membership function, which assigns a value (called a membership grade) of between 0 and 1 to every element in X. Z^*shall denote the (classic) set of all such pairs. For any set A, μ_A denotes the membership function of A, and A's cardinality (noted $|A|$) is just the sum of the membership grades of all its elements. Now, $\forall x \in X$, x is called

- **fully accessible** in the (fuzzy) set A if $\mu_A(x) = 1$;
- **not available** in the (fuzzy) set A if $\mu_A(x) = 0$;
- **partially accessible** in the (fuzzy) set A if $0 < \mu_A(x) < 1$.

Representing opportunity sets as fuzzy sets is extremely natural as an extension of the approaches discussed in Section I. The concept of a fuzzy set is simply a generalization of the concept of a set—classic sets are degenerate fuzzy sets in which all elements are assumed to have a membership grade of either 1 or 0. Fuzzy sets retain the concept of cardinality, and all classic set operators have fuzzy analogues; the only difference is that the device of fuzzy sets allows us to explicitly represent the idea that options admit of degrees of availability. Intuitively, this is just what membership grades denote.

It is an attractive feature of this approach that we are free to interpret the membership grades in different (not necessarily competing) ways. For example, we can interpret them as denoting distances from possible worlds: the larger an option's membership grade, the further away the nearest possible world in which it is not available. This interpretation would fit republican or independence-based conceptions of freedom, such as those defended by Pettit or List and Valentini,[53] according to which an individual is free to φ if and only if there is a robust absence of externally imposed constraints on their φ-ing (i.e., the nearest possible world in which constraints are imposed is quite distant). Following my arguments in Section II, however, I propose to interpret membership grades

53. Philip Pettit, *A Theory of Freedom: From the Psychology to the Politics of Agency* (New York: Oxford University Press, 2001); Christian List and Laura Valentini, "Freedom as Independence," *Ethics* 126 (2016): 1043–74.

as conditional probabilities. In other words, I interpret $m(x)$ as the chance that an agent can successfully choose x if they intend to.

Obviously, the value of $m(x)$ will depend on several factors; agents always face particular constraints (physical and mental limitations, poverty, lack of skills or education, geographic isolation, coercion, etc.) that simply rule out particular options, and as we saw earlier, there are other factors which conspire to make options less accessible (e.g., whether these options require the recruitment of motivation to be chosen, how much motivation is required, how strong-willed the agent in question is, what other options are available, and how salient each option is), without ruling them out entirely. Accordingly, a set's membership function is actually an n-argument function, where each of the n arguments denotes one factor that impacts option accessibility.

Note that the value of an option's membership grade need not track very closely how effortful it is to choose that option: if you are iron-willed and you always succeed in choosing as you resolve to or as you judge you ought, even when faced with such temptation as requires strenuous effort from you to resist, then my approach might count all your options as fully accessible. This seems right: if you can always overcome temptation when you want to, then temptation is no constraint on your freedom. You may prefer to be rid of temptation, but then being exposed to temptation simply means that the level of welfare you achieve is lower than it might otherwise have been, not that you are any less capable of doing, being, or choosing anything. Likewise, if you are incredibly weak-willed and abandon your resolutions as soon as it becomes hard to maintain them, then the membership grade of some of your options will be very low. Again, this seems right.

Observe also that while it is the problem of weakness of will that motivates my proposal, my framework is perfectly general. It imposes no constraint on the list of factors that can be included as arguments of a set's membership function, nor does it require the inclusion of any particular factor, and it allows for flexibility in the way that we can represent the impact of these factors that we do include on people's freedom. Notably, my framework permits us to include external constraints among the arguments of a set's membership function and to represent the effect of these constraints as that of merely diminishing the accessibility of an option. Remember that because standard approaches to the measurement of freedom assume bivalence in set membership, it is impossible for a constraint on one's freedom to be a constraint unless it entirely rules out certain options. But consider the following case devised by Michael Garnett: you wake up in a room, and you find that the door is locked by a padlock whose combination you do not know.[54] Is the option to leave the room

54. Michael Garnett, "Ignorance, Incompetence, and the Concept of Liberty," *Journal of Political Philosophy* 15 (2007): 428–46.

available to you? Intuitively, it is hard to say: on the one hand, all you need to do to get out is put in the right combination, but on the other hand, if all you can do is guess at possible combinations, it looks like you may never get out. My account deals with this case rather elegantly: the option to get out is available to you, but it is not very accessible, because, conditional on your intending to put in the right combination (whatever it is) and get out, the chance that you will do so is vanishingly small. And naturally, if you are somehow able to narrow down the list of possible combinations, then my analysis implies that your freedom has just increased substantially, since the chance that you will get out soon, given that you intend to, has just increased.

This analysis of probabilistic external constraints is interestingly distinct from other analyses that have been proposed. Carter also argues that a measure of freedom must incorporate probabilistic judgments in some way, but his proposal is that we identify an individual's degree of freedom with her expected number of options.[55] The expected number of options available to an individual is simply the sum of all possible options, discounted by the unconditional probability of there being any externally imposed preventing conditions on the option which makes it fully unavailable. Importantly, this proposal retains bivalence, in that for any option x and any opportunity set A, either x is fully in A or x is not in A; it is just that there is some uncertainty over which A will in fact be yours. This proposal has substantively different implications than mine, for the simple reason that the presence or absence of a preventing condition on my φ-ing need not be probabilistically independent of my intention to φ.

Imagine that I have a sworn enemy who is a mind reader and who will attempt to prevent me from carrying out whatever intention he reads in me. In that case, the conditional probability that there will be a constraint on my φ-ing given that I intend to φ will be much higher than the conditional probability that there will be a constraint on my φ-ing given that I do not intend to φ. So there being a constraint on my φ-ing is not probabilistically independent of my intending to φ. In this situation, for any φ, the chance that I will φ, given that I intend to, will be very low, since there is in that case a high chance of a preventing condition. In contrast, for any φ, the unconditional probability of there being a preventing condition on my φ-ing may be quite low, since my enemy won't bother to impose constraints on my φ-ing if he does not discern in me any intention to φ. It is only if I intend to φ that the chance of there being a constraint on my φ-ing increases. Carter's proposal would imply in this case that I enjoy a good deal of freedom, almost as much, in fact, as if I had no enemy bent on thwarting my will; my proposal, however, implies that I enjoy very little freedom, in the sense that anyone with a very small number of fully available options would enjoy greater freedom than I. And this, in fact, seems

55. Carter, *Measure of Freedom*, 191.

right. So not only is my analysis of probabilistic external constraints different from Carter's, but it also has more plausible implications.[56]

Most importantly, however, my framework also allows us to refine the standard ranking rules surveyed in Section I in ways that are sensitive to the heterogeneity in the capacity of individuals to carry out their choices. To see this, just suppose that in the case of Willy and Regina, all four of Regina's options are fully accessible, but Willy only has one fully accessible option (going out to drink), with his three work options being only partially accessible. Letting $\succsim$ be defined over Z^*, the fuzzy analogue to the simple cardinality rule now tells us that for any two opportunity sets A and B belonging to Z^*, $A \succsim B$ just in case $|A| \geq |B|$.

Let A and B denote, respectively, Regina's and Willy's opportunity sets. Whereas the simple cardinality rule would tell us that $A \sim B$, since Willy and Regina both have four options available, the fuzzy cardinality rule tells us that $A \succ B$, because $|A| = 4 > |B| > 1$. And if we assume that by depriving Willy of the choice to go out drinking we increase the membership grade of his three work options, the fuzzy cardinality rule tells us that Willy's freedom increases provided that the change in the sum of the membership grades of the three work options is greater than 1. This rule will also rank Cassandra's opportunity set as offering strictly more freedom than Perseus's. Thus, the fuzzy cardinality rule generates more intuitively correct rankings than the simple cardinality rule.

Any other ranking rule can similarly be generalized to a ranking rule over Z^*. It is easy, for instance, to incorporate a minimal desirability constraint on the formula above, by stipulatively assigning a membership grade of 0 to any option which fails to satisfy the desirability constraint. Considerations of similarity can be incorporated in a similar fashion. An extremely simple proposal might run as follows: instead of counting how many (minimally desirable) options in an opportunity set, we could count how many (minimally desirable) option types there are and identify the membership grade of each option type with the membership grade of the most accessible option token of that type.

There is much to be said in favor of my proposal, then, to represent opportunity sets as fuzzy sets. It is a natural generalization of the existing approaches to the measurement of freedom, but the great advantage of

56. Matthew Kramer (*Quality of Freedom*, 175) also argues that the ascription to an agent of the freedom to φ must carry "a probabilistic qualification" indicating what the chances are that the agent will enjoy that freedom. However, he does not state explicitly how chances are to be interpreted, or what form the qualification is to have. That being said, since he is a proponent of pure negative freedom, like Carter, and believes that one is unfree to φ just in case other agents impose preventing conditions on one's φ-ing that make it impossible for one to φ, one can reasonably impute to him the same view as Carter regarding how the uncertainty of there being constraints on one's actions should affect the measure of one's freedom. To the extent, then, that Kramer and Carter agree with each other, my view is also in disagreement with Kramer's.

this approach is that it allows us to take account of the fact that options may be more or less accessible. As demonstrated, it allows us to reach more accurate measures of individual freedom.

VI. APPLICATION

One of the main motivations for developing a measure of freedom is that the concept of freedom is central to many debates in political philosophy and public policy. In this section, I briefly explore how one might try to use my framework to generate new insights into old questions by applying it to the ethics of nudge.

Nudging consists in changing the way in which choices are presented to people so as to induce them to make choices that are better for them, or that are thought to be better for society. Crucially, nudges operate by exploiting people's reflexive choice habits, their unconscious, irrational tendencies to choose in particular ways when choices are presented to them one way rather than another.[57] No steps are taken to remove any options from individuals, nor to impose any burdens on individuals who rationally choose the behaviors we are aiming to discourage, nor to provide them with new information that might cause them to consciously revise their behaviors (though nudging may supplement information campaigns).[58]

The no-burden requirement is a bit ambiguous, but what Thaler and Sunstein have in mind is that we should not make it more expensive or more time-consuming to engage in the sort of behavior we're trying to discourage, because this would make it impossible for individuals to engage in it and spend this money or time on other things.[59] In a way, then, the no-burden requirement can be interpreted as reemphasizing the crucial point that nudges must not make any previously available option unavailable, where options are conceived here, as before, as conjunctively possible courses of action.

A paradigm case of nudging is Save More Tomorrow, which was a program proposed by Thaler and Benartzi to increase employee's contributions to their 401(k) retirement plans.[60] Under this plan, employees

57. Luc Bovens, "The Ethics of Nudge," in *Preference Change: Approaches from Philosophy, Economics and Psychology*, ed. Till Grüne-Yanoff and Sven Ove Hansson (Dordrecht: Springer, 2009), 207–19.

58. Thaler and Sunstein actually count providing individuals with information as nudging them, but Bovens ("Ethics of Nudge") and Hausman and Welch argue that this is a mistake. Richard H. Thaler and Cass Sunstein, "Libertarian Paternalism," *American Economic Review* 90 (2003): 175–79; Cass Sunstein, *On Freedom* (Princeton, NJ: Princeton University Press 2019); Dan Hausman and Brynn Welch, "Debate: To Nudge or Not to Nudge," *Journal of Political Philosophy* 18 (2010): 123–36.

59. Thaler and Sunstein, "Libertarian Paternalism."

60. Richard H. Thaler and Shlomo Benartzi, "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy* 112 (2003): S164–87.

were asked some time before receiving their next pay raise whether or not they wanted to commit this raise to their pension plan, as opposed to being asked once they received their raise. This plan tends to greatly increase savings rates: in the first company to participate, savings went from 3.5 percent to 11.6 percent. As Bovens explains, Save More Tomorrow exploits two design flaws in human psychology: first, people find it harder to part with what they already have than with what they do not yet have (i.e., the endowment effect); second, people find it harder to resist temptation than to make provisions for the future so that they are not tempted at all. Employees do remain, of course, free to choose either way. So, although nudging is paternalistic, to the extent that we are attempting to guide people's behavior in ways that we judge to be best for them, Thaler and Sunstein argue that it is a benign, "libertarian" form of paternalism because it does not reduce the available range of choices.[61]

Thaler and Sunstein's argument has proven controversial. Hausman and Welch argue that the "libertarian" credentials of nudges are dubious, because although nudges do not remove options, they may undermine individual control over their choices and evaluations, by making individuals act in ways that reflect the social engineer's designs, not the individual's.[62] Bovens, likewise, though he supports some nudges, is more circumspect than Thaler and Sunstein, arguing that nudging is problematic if it aims to make us choose in ways that are not in line with our actual preferences.[63] He gives the example that society may be better off if I am nudged not to place an additional fishing boat in already overfished waters, but that I may be worse off for being so nudged. This case raises some concerns, he believes, because we are being nudged to choose in ways that are aberrant, out of touch with our overall judgments of what is in our interest.[64] In contrast, he argues that a nudge is much less worrisome if it brings our agency into better alignment with our preferences; notably, in cases where we are limited by ignorance, weakness of will (I will come back to this below), status quo bias, or some irrational queasiness from making the choice that best fits our preferences, nudging may induce us to choose in the way that, on reflection, we judge is best.

My framework offers a rather new perspective on this debate. As I will show, it is not in general true that nudging does not affect the degree of individuals' freedom. In particular, it is certainly false to claim that nudging never reduces individual freedom. More surprisingly, however, my framework implies that in some cases nudging may actually expand

61. Thaler and Sunstein, "Libertarian Paternalism."
62. Hausman and Welch, "Debate."
63. Bovens, "Ethics of Nudge," 218.
64. Ibid., 213.

individual freedom. Indeed, the most attractive nudges will be those that do expand individual freedom.

Consider first Thaler and Sunstein's response to the objection that nudging is paternalistic. If the main reason that paternalism is objectionable is that it curtails the freedom of individuals against their consent,[65] then it won't do to point out that nudging does not restrict the range of choices available to agents: as we have seen, this is no guarantee that nudging does not restrict individual freedom. Consider, for instance, the "Don't Mess with Texas" campaign, which aimed to reduce littering by saturating road signs and advertisements in radio and in print with the phrase "Don't Mess with Texas," often featuring popular celebrities like Stevie Ray Vaughan in their ads to drive the slogan home. As Hausman and Welch point out, although the campaign was informational (it ran and continues to run educational campaigns to teach Texans about the harms of littering), its central slogan and messaging "attempted to create a machismo image for those who don't litter," in essence exploiting Texans' self-image to make them want to not litter.[66] As any standard model of behavior in the economics of identity would predict, the effect of successfully portraying some behavior as being prescribed by individuals' conception of their identity will be to generate a desire to engage in it.[67] This desire, of course, may run contrary to one's better judgment or to one's resolutions, and for this reason it may diminish one's freedom of choice if motivation needs to be recruited to overcome it.[68]

What this example shows is that Thaler and Sunstein's "no substantial burden" requirement does not suffice to guarantee that nudging does not restrict freedom. Rather, two conditions must be met: (1) that no option be ruled out, and (2) that no option be made less accessible. These conditions are not met in the case of the "Don't Mess with Texas" campaign. But if both these conditions are met, then my approach to the measurement of freedom implies that individual freedom is not restricted by nudging.[69] The charge of paternalism thus loses some of its sting, as Thaler and Sunstein claim, though in fewer cases than they hoped.

65. Richard Arneson, "Mill versus Paternalism," *Ethics* 90 (1980): 470–89.

66. Hausman and Welch, "Debate," 134.

67. George Akerlof and Rachel Kranton, "Economics and Identity," *Quarterly Journal of Economics* 115 (2000): 715–53. More precisely, in Akerlof and Kranton's terms, individuals will suffer a loss of identity-related utility if they behave against the behavioral prescriptions which they take to follow from having the identity they ascribe to themselves—e.g., "real Texans don't litter."

68. Perhaps this example is not so troubling, as the aim was only to reduce littering, but suppose instead that it had been to induce Texans to mass-purchase useless and expensive consumer goods.

69. In some cases, nudging might decrease the accessibility of some options and yet also increase the availability of others; for instance, by making unhealthy food choices in a cafeteria less salient, I may reduce the force of temptation they exert on dieters, but obviously

More surprisingly, though, and more interestingly, my framework implies that in some cases nudging may actually expand individual freedom. Here again we turn to cases of weakness of will. In Save More Tomorrow, the reason individuals are more prudent if asked before rather than after receiving their raise is that at t_{after} they are exposed to a source temptation (i.e., the check in their hands), but not at t_{before}, and so they more competently manage their spending decisions at t_{before}. The nudge, in this case, consists in nothing more than removing a source of temptation from decision-makers by changing the time at which they must make their decision, thus making the option to save their next raise toward retirement more accessible, without in the interim making it any harder for them to spend their money in other ways or changing how they may choose to spend it. In this case, a fuzzied cardinality-based ranking implies that nudging will expand individual freedom of choice, and this should take the sting out of the charge of paternalism. Given, then, that nudges which are limited to removing sources of temptation increase individual freedom, these will be the most attractive sorts of nudges.[70]

I should note that I am not the first to argue that nudging may expand freedom in cases of weakness of will. Sunstein also defends this claim.[71] However, on his view, it is because nudges allow individuals to achieve outcomes they prefer that they may expand individual freedom. Plainly, this explanation commits Sunstein to a strongly value-laden measure of freedom. Being free, Sunstein tells us, is (in part) a matter of actually achieving preferred outcomes. What is distinctive about my claim is precisely that it is couched in a value-neutral account of what it means to be more or less free to pursue a given course of action. And this, I submit, makes my conclusion more surprising than Sunstein's: it is not news that exploiting certain features of people's psychology may make them better off—that much was already clear from examples like Save More Tomorrow—but it is somewhat surprising that this could make them more capable agents.

In sum, by looking at the ethics of nudge through the lens of my framework, we are better able to judge when nudging is compatible with

---

this will increase the salience of healthy food items, and if people have an irrational tendency to pick the most salient options, this could decrease the accessibility of the unhealthy options. In this case, the overall effect on freedom could be a wash, though strictly speaking condition 2 is not satisfied. We can perhaps reformulate it as the condition that no option be made less accessible without compensating increases in the accessibility of other options.

70. A potentially more chilling implication of my approach is that subliminal messaging could also be freedom enhancing, if the subliminal messages were designed to manufacture desires that helped us overcome weakness of will. Perhaps implanting in people a desire to be resolute would make them better at resisting temptation.

71. Sunstein, *On Freedom*, 63.

freedom. More could be said by looking at additional cases, but I hope this brief discussion has shown why my framework might be of interest in real-world cases.

VII. CONCLUSION

Cases of weakness of will reveal that measuring individual freedom requires us to take account of the heterogeneity in the accessibility of options. Because the assumption of bivalence is baked into extant measures of freedom, they lack the machinery with which to take this heterogeneity seriously. Thus, otherwise attractive ranking principles generate bizarre rankings in cases where weakness of will gets involved.

My proposal resolves this problem. Representing opportunity sets as fuzzy sets allows us to model the fact that option accessibility admits of degrees. And in redefining the domain of the "offers at least as much freedom as" relation to the set of all "fuzzied" opportunity sets, we can generalize the standard models so as to both preserve what is attractive about our preferred ranking rules and generate the intuitively correct rankings in the problem cases involving weakness of will. This framework has the advantage of being at once more precise in its treatment of how weakness of will constrains human freedom than other accounts while remaining perfectly general, providing an attractive analysis of constraints that applies equally well to internal constraints as to external ones. Finally, this framework's appeal also lies in its potential for new insights into other problems, as illustrated in the case of the ethics of nudge.