Value Capture

<anonymized for submission>

Abstract

Value capture occurs when an agent enters a social environment which presents external expressions of value — which are often simplified, standardized, and quantified — and those external versions come to dominate our reasoning and motivations. Examples include becoming motivated by Twitter Likes and Retweets, citation rates, ranked lists of best schools, and Grade Point Averages. We are vulnerable to value capture because of the competitive advantage that such pre-packaged value expressions have in our reasoning and our communications. But when we internalize such metrics, we damage our own autonomy. In value capture, we outsource the process of deliberating on our values. And that outsourcing cuts off one of the key benefits of personal deliberation. When we tailor our values to ourselves, we can fine-tune them to fit our own particular psychology and place in the world. But in value capture, we buy our values off the rack.

Body text

A friend told me a story about how technology can change people. She and her husband had planned a European vacation with another retired couple. She'd had high hopes for the vacation — for rekindling old friendships through communal exploration and shared conversation. But these hopes, she said, were dashed by the other couple's relationship to their FitBits. Her friends' daily decisions were ruled by their attempts to maximize their step-counts. Many nights, she'd try to arrange some group dinner only to have her friends reply, "Sorry, we can't make it tonight; we haven't met our step-goals yet."

The magnetism of that measure may seem rather insidious. Other values seem to have fallen by the wayside. They might have considered the artistic and historic value of their destination, the joys of leisurely exploration, or the loveliness of a shared meal with friends. Even if their main interest was exercise, there are still plenty of other values and considerations that could come into play. What kind of pace and terrain is most pleasant? What feels best for body and mind? But such varied considerations seem to have fallen by the wayside.

As a first pass, here's what I think happened. It's not that this couple consciously decided that these other things didn't matter. It's that FitBit's evaluative clarity and pervasiveness somehow swamped those other considerations. Numbers can speak loudly in our thinking, especially when they are tied into clear and pervasive reporting systems. And it is hard to quantify the value of seeing historical artifacts. There is no clear leaderboard which ranks our daily achievements in having "pleasant meals with friends". A FitBit, on the other hand, offers a ready-made measure for success: step-counts. That measure permits easy ranking, and a satisfyingly quantized sense of progress. And the physical artifact — the watch itself — constantly spotlights that measure. A FitBit is always on your wrist, always eager to display your step-count. And its outputs are highly portable. Step-counts can easily be exported into other formats, like a fitness-tracking spreadsheet. The step-count measure has a competitive advantage in the modern ecosystem of technologized self-evaluation.

And it's not just FitBit. We see the motivational magnetism of quantification across so many other domains. Students start their educational lives full of curiosity and wonder, and emerge obsessed with their grades. Bright-eyed activists start their pre-law education dedicated to the cause. Four years later, they're laser-focused on the *US News & World Report*'s law school rankings. Young philosophers enter grad school entranced by philosophy's depths and profundities; they emerge obsessed with citation rates, impact factors, and their institution's ranking in the Philosophical Gourmand Report. We get on Twitter hoping for conversation and connection, but find ourselves caring about getting more Likes, Retweets, and Followers.

Let's call this phenomenon *value capture.*[1] Value captures happens when:

1. An agent encounters a social environment that presents an external and explicit expression of some value — which is often simplified, standardized, and/or quantified.

2. The external expression of value comes to dominate the agent's practical reasoning in the relevant domain.

The most obvious case of value capture involves an individual's values being swamped by the values presented by some large-scale group. Some corporation, institution, or community generates an explicit expression of value, and that expression gets internalized by an individual. In institutional settings, such values are often highly explicated, and

1 The term "value capture" is inspired by the term "regulatory capture", which occurs when government regulators' interests become more aligned with the interests of those who they are supposed to be regulating, rather than with the public interest.

packaged with fixed and standardized criteria for application. I will focus on the most vivid cases of value capture: capture by a quantified expressions of value — including intentional gamifications like Twitter's scoring systems, measures like citation rates, and rankings like the *US News & World Report*'s law school report. But value capture can also happen through explicit, qualitative representations of values — like university mission statements; the standardized language of justification common to institutional settings; and other moral cliches.2

I'm concerned here, not with where we source our values from, but what we do with them once we have them. The worry here is not one of origin, but of adaptation. Suppose I acquire the starting seed of a value from some external source, but then adjust it fit my particular personality, my larger value network, and my particular life situation. That isn't value capture, because the externally-sourced version isn't dominant. What's dominant is some variant — a version which I've significantly tailored to my own needs. My worry is about when we fix our values to some particular external presentation — when we outsource the process of deliberating about and adjusting our values. Value capture is the phenomenon of ingesting, wholesale, an externally-sourced value, and letting it dominate one's practical

---

2 A notable example from education is Bloom's Hierarchy, which classifies student cognitive achievement into levels, and provides terms to go with each level. The hierarchy has come to largely serve as a fairly meaningless, but regularized, set of terms to describe the increasing levels of student understanding. When I was handed Bloom's Hierarchy by a colleague, I was told, "Nobody actually knows what these words mean, but if you put them in your reports, administrators will nod and approve your funding requests."

reasoning in its original form.

When I say that those external values "come to dominate," I mean this in a broad sense. The purest cases of value capture are those where an externally-sourced value comes to wholly replace or displace other values. But more common, I suspect, are cases where an external expression comes to guide the application of some value one already holds, without fully replacing the original value. Instead, the external expression comes to dominate how a person articulates and applies their value during practical deliberation. It controls the effective expression of that value. For example, most Twitter users probably don't think that the Retweet and Like numbers capture the full range of their communicative values. They would, if asked, likely tell you that they used Twitter to connect to people, to learn about the world, to enjoy themselves. But if the pursuit of Retweets and Likes comes to guide their choice of topics, content, and tone in their tweets, then they have been value captured. Similarly, I suspect that the average FitBit user would say that their primary interest is in health and fitness. They are aware, on reflection, that step counts are only a simplified proxy for health or fitness. But insofar as the pursuit of step counts dominates their daily decisions regarding exercise, then they have been value captured. Call such cases *value articulation capture*. And, since "dominance" is a spectrum concept, rather than a binary one, then value capture is also a spectrum concept. One can be dominated to varying degrees. This is particularly clearly in articulation capture, where a particular articulation like step-counts can guide a many, but not all, of one's decisions in the domain.

Value articulation capture is especially powerful when it guides self-evaluation. Say that I, an academic, care about the pursuit of truth. But suppose that whenever I try to evaluate the success of one my articles, I turn to certain metrics — like the citation rate or the status of the publication venue on some ranked list. And when I evaluate my overall success as an academic, I turn to metrics like my total citation rate or the status of my institution on some ranked list. In that case, it is those institutional metrics, and not the values I report upon reflection, which effectively dominate my self-evaluation. Here, the metric gains dominance by capturing, not my reflective expression of my core values, but the effective criteria with which I perform self-evaluations. It sets the specific terms in which I determine whether I have fulfilled my core values. (From here on out, I will speak of "values" being captured, for brevity's sake — but I mean both pure value capture and value articulation capture.3)

I am interested in understanding a very specific relationship we can have with our values, that seems particularly characteristic of modern institutional life. I mean to be picking out those cases in which an external value is absorbed and deployed *without further adjustment.* We use the external value just as we found it, without further contouring it, interpreting it, or fine-tuning it to ourselves. I am not trying to defend some radical notion of individuality, where we must all invent our own values out of nothing, and maintain them in some solitary purity. Our values grow from common soil. We share them, we adjust them in light of our relationships with other people, we form communities around them. But in many of those cases, we also have some ability to negotiate the details of those values. We are partial authors of those values, or we can interpret some common concept in our own particular

3 I am also be happy to take onboard the view that one's core values were whatever it was that guided one's actions and deliberations and not how one conceived of or explicated them. The same range of cases would still count as value capture, though the internal taxonomy would shift.

way. I may participate in an artistic community that jointly pursues the creation of, say, funny improv comedy — but each of us might have our own personal spin on the notion of "funny". Value capture occurs when we ingest our values wholesale from external sources, leaving it to those external sources to fix the precise details and contours of those values. I am interested in the use of pre-packaged values.

I mean "agents" here in a broad sense. A case which leaps readily to mind is that of the lone individual coming to internalize some bureaucracy's measures and metrics. But I also intend the notion of "value capture" to include cases in which one group's shared values are swamped by another group's expressed values — often a much larger or more prominent one. For example, a small humanities department could be value-captured by their university's barrage of measures, focused on student satisfaction surveys and student graduation speed. Or a particular university might be value-captured by, say, some external ranking system, published in some magazine. However, the case of group-level value captured is significantly more complex, and depends on resolving some controversies about the nature of shared values and collective intention. As a first step, I'll focus on working out the details of the case of individuals being value captured — though I will attempt to broaden and complicate the picture towards the end.

So let's start by focusing on cases of value capture of individuals, by the value presentations of large groups. My two central questions are: why are we so attracted to these pre-packaged values? And how might internalizing them harm us? The answers will turn out to be deeply linked. What makes external values so seductive is precisely what makes them so harmful to ingest. Take the starkest case: value capture by quantified institutional metrics. Such metrics have been engineered to be simple, clear and easy to use. Their application is carefully standardized, to provide for the swift and frictionless communication of evaluations. So when we take them on, we align our values — making ourselves easier to understand and our justifications easier to uptake. But the system is frictionless only when people all adhere to the same system. Easy communication depends on that system remaining relatively stable — which, in turn, requires that it resist individual tailoring.

The problem of value capture is not the same as the problem of perverse incentives. Perverse incentives are quite familiar. Students are incentivized to get a good GPA, in order to get a job. Academics are incentivized to maximize their publication rate and status for hiring and promotion. These incentives can undercut the real purposes of their activities: education and scholarship. However, perverse incentives usually work by leveraging some instrumental resource, like money. Value capture concerns the next step of intrusiveness — when individuals internalize those external values. Value capture happens when students come to pursue GPA as important in and of itself; when academics come to treat citation rates and publication status as their primary goal. And I want to think about the special harms that come from that internalization. Something crucial is lost when a person becomes thoroughly value captured — when their senses of meaning, purpose, and value are adopted wholesale from external sources.

Many of my examples will concern quantified metrics. My goal here is not to criticize quantification in general, nor is it to question the validity of well-conducted science. Quantification can be invaluable, when used in the appropriate circumstances, and when the measures are constructed with appropriate sensitivity to the task at hand. But the typical pressures which drive the creation of institutional metrics and measures often leave them insensitive to the subtler features of our experience. I am particularly concerned about the attempts, by large-scale institutions, to quantify the fundamental terms by which we evaluate our own lives — to create metrics of happiness, education, health, and well-being. First, the social forces which drive the generation of quantified metrics often prefer simple, easily measured qualities. This is not a bad thing, per se. We need institutions, and institutions need to simplify their view of individuals in order to manage the information overload. But when we internalize those values, we will adopt their institutional oversimplifications — and take onboard values which ignore the more complex and variable features of our experience. When we internalize such metrics, we are internalizing the institution's-eye view of ourselves.

And: irregardless of whether such external values are quantified or oversimplified, when we internalize external values, we undercut our participation in the process of value self-determination. The existence of a pervasive, pre-packaged metric for happiness invites us skip out on the process of value self-determination. In so doing, we are stepping back from the process of carefully adjusting their values to our own particular personality and place in the world. We are putting the formulation of our values into the hands of an external source. There is a complex trade-off here. Such outsourcing offers us certain goods — like harmony and cognitive ease — at the price of autonomous control over the details of our values. And, in some cases, the goods that arise from precise group coordination may outweigh the goods of autonomous, fine-grained self-governance. If, for example, there were a good gamification to encourage actions that might avert climate change, or control the COVID-19 pandemic, we should almost certainly deploy it. But in other cases, the benefits are marginal and the sacrifice substantial. And we often take the trade without being entirely aware of what we are paying. I am not saying that we should never outsource — but I think we should be more fully aware of the cost. The more of our core values we outsource, the more we chip away at our substantive authorship over our own life-goals. And some forms of valuing are particularly costly to outsource, because they drive our action in domains that are particularly personal — whose importance is deeply connected to the rich and subtle phenomena of our own particular life-experience. I'm thinking here of the goods of aesthetic experience, of career satisfaction, of fitness and of family life. Many goods are most available to us when we have, if not bespoke values, then, at least, significantly custom-tailored ones. But with value capture, we buy our values off-the-rack.

Value capture and autonomy

Here's a particularly clear example of value capture and its consequences. Sociologists Wendy Espeland and Michael Sauder document, in *Engines of Anxiety*, their ten-year investigation into the effects of the *US News and World Report*'s law school rankings on legal education and legal culture. Before the *USN&WR*, there were no law school rankings. Students often picked law schools through a complex and personalized process of evaluation. They got to know a school by reading about its mission, by talking to people, or by visiting. In the process, many students had to figure out what they wanted from a legal education. Some students might care about the research prowess of the faculty; others about the school's connections to high-paying corporate firms; others about the school's deep support for social activism (Espeland and Sauder, 2016).

The rankings displaced all that. Once the *USN&WR* started publishing its rankings, they came to dominate the choice process for most students. Espeland and Sauder monitored online discussions between prospective law student; they found that the *USN&WR* ranking had become the most important basis for judging school desirability. And the same is true for non-students: the public perception of law schools immediately re-oriented itself along the *USN&WR*'s rankings. Law schools, they say, once pursued a far more varied set of missions. Some emphasized admitting students from underprivileged backgrounds; others sought students interested in social activism. But following such a distinctive mission invariably meant dropping spots in the rankings — which promptly resulted in precipitous drops in donations and student interest. Most schools, report Espeland and Sauder, have since abandoned their original missions. They have re-oriented their admissions and educational procedures towards performance in the *USN&WR*'s ranking calculations — which heavily emphasize measures such as the GPA and LSAT score of the incoming class, and the employment rate of the outgoing class (43).

To put their analysis in philosophers' terms: before the rankings, law students and law schools could pursue a complex plurality of values. And they could settle on those differing values with a fair amount of independence. Law students could decide for themselves what they cared about in a legal education, and law schools had some leeway in setting what sort of legal education they wished to provide. The rankings ended that robust plurality, ironing out the variance and independence, and substituting for them the forced linearity of a single ranking system.

In the case of the law schools themselves, the problem may simply be one of perverse incentives. [4] But with prospective law students, the problem runs much deeper. The

---

[4] Espalund and Sauder (2016) chart some cases of direct value change, but also chart

presence of the rankings seems to exert a magnetic pull over what prospective students value. Some students, of course, are merely responding to incentives — since potential employers also care about law school's rankings. But a majority of students, say Espeland and Sauder, seem just to care directly about those rankings. Instead of exploring their own values and desires for their legal education, they seem to presume that the process of going to law school should be oriented towards getting into the "best" law school, where "best" is determined strictly by the rankings. The existence of that clear, vivid, objective-seeming list offers an easy substitute for the process of personal value deliberation.

What is the harm in value capture? Espeland and Sauder's description suggests many possibilities. I will focus for the remainder of this paper on one thread. Value capture, I suggest, significantly disrupts the process of value self-determination. We could be developing our values for ourselves, tailoring them to our own psychologies, needs, and to our particular place in the world. Value capture cuts off that process, substituting pre-fabricated and standardized values for self-tailored ones. As anthropologist Sally Engle Merry puts it, the culture of indicators and metrics is "a form of governance that engages a person in governing himself or herself in terms of standards set by others" (Merry, 2016,

significant institutional churn. That is, in many cases, the rankings place such an enormous amount of uncomfortable pressure on the system, that adherents to the pre-ranking value system eventually become disenchanted and leave administration. They are typically replaced with new administrators who take improved placement on the rankings to be the primary goal of law school administrators. This institutional churn is interestingly different from the process of the value capture of individuals that I describe in this paper.

33).5

In searching for a fitness routine, I could determine the value of fitness activities for myself. I could consider whether I wanted to pursue the zen-like trance states of long-distance running, or the sharply intellectual pleasures of solving bouldering problems in rock climbing, or the intense full-body effort of heavy deadlifting. But when I am value

---

5 Merry here is explicitly developing Nikolas Rose's analysis of "government at a distance" — the attempts of governments to govern individuals by governing their souls (Rose, 1999). Though my analysis here obviously bears many resemblances to, and influences from, Rose's analysis. I have avoided relying extensively on Rose's discussion here, because Rose makes some theoretical assumptions that may be controversial, especially to analytic philosophers. I do not take all instances of value capture to be the result of intentional attempts at governance through intrusion into values. Value capture also includes many cases where metrics are created for information-processing purposes, and then are internalized by people because of their seductive clarity — even when they have not been made for the purpose of that seduction. In this way, this analysis differs from many neo-Foucaultian accounts which emphasize the intentionally manipulative aspect. Many of the sources on which I draw - Porter, Scott, Merry, Espeland and Sauder, and Bowker and Starr — are sociologists and anthropologists working very much in the Foucaultian tradition. But their work — grounded as it is in careful empirical work — reveals a much more complicated relationship between institutional function and seductive effect. Their discussions involve a thinner set of presuppositions, and their arguments are more spiritually amenable to the average analytic philosopher.

captured by FitBit, I outsource the process of value deliberation to the technology, and to the institutions behind it. This leads to some distinctive harms. When we are value captured, we no longer substantially participate in tailoring those values. And, accordingly, the values we end up with won't be tailored to our particular selves. This suggests two angles from which to think about the harms of value capture. First, value capture damages our *autonomy*, by taking the formulation of our values out of our hands. Insofar as autonomy about the formulation of one's own values is intrinsically valuable, then this loss is intrinsically harmful. Second, the loss of autonomy over our values has a further, instrumental harm. It decreases the *fittingness* of our values, by making the formulation of our values responsive, not to our own interests, psychology, and experience, but to the interests of large-scale institutions.

These claims about autonomy might strike some as non-starters. After all, autonomy is self-rule. We are autonomous so long as we choose which rules to follow. And voluntarily choosing to take an external measure and make it one's own value is surely an instance of self-rule. It is literally giving oneself a rule by which to measure oneself. What's more, many of these voluntary self-gamification are done for the sake of increasing one's willpower. Many people use Fitbits in an intentional effort to increase their motivation for exercise. Gamifications are often sold as a technology for overcoming weakness of will (McGonigal, 2011; Walz et. al. 2015). And what could be a better instance of self-rule than voluntarily adopting a rule to amplify one's willpower?6

---

6 I am meaning to indicate here a tradition of thinking about autonomy in terms of overcoming weakness of the will, which traces some of its lineage to Jon Eslter's discussion of

The initial decision to outsource may be a voluntary act. But what I am outsourcing is, in fact, a fundamental part of my existence as a continuously and finely self-determining agency. I am outsourcing my capacity to further tailor and adjust my values, in response to my experiences of living in the world. The initial act may be autonomous, but it damages my autonomy going forward. To voluntarily engage in value capture is to voluntarily renounce one's control over the detailed content of one's own values.

Value capture does offer a number of benefits: it saves cognitive resources; it fits one's values into well-machined structures of justification and information processing. But its benefits are inextricably entangled with its costs. Value capture works to ease our burden of justification and self-exploration, precisely by letting us skip doing the value tailoring for ourselves. Value capture can plug us into a carefully manufactured system of amplified motivation — like the way FitBit offers us an increase motivation by drip-feeding us a steady tickle of points (McGonigal, 2011). But that motivational boost depends on actually caring about the points, as they are presented by the technology. We get that motivational boost only when we align our values with FitBit's measures. And you cannot tinker with the way a FitBit evaluates you. A FitBit, one might say, offers a pre-fabricated add-on to the will. It offers an increase in willpower, but only towards those pre-set targets.

Some value capture may happens on the sly, without the conscious and voluntary participation of the value-captured person. Such covert value capture would undermine autonomy; that much seems uncontroversial. But many cases of value capture involve willing participation — as with the voluntary adoption of a gamified technology. The typical story, I

---

self-constraint and strength of will (Elster, 1998).

think, is that a FitBit user knows that when they wear a FitBit regularly, their motivations and reasoning are swamped by the FitBit's measures — and decides, of their own accord, to keep doing so. I wish to show that in those cases, too, a person damages their own autonomy.

How is it possible to autonomously undermine one's own autonomy? First, let me be clear about the nature of my target. As Nomy Arpaly points out, philosophers often conflate very different senses of autonomy. In one sense, says Arpaly, "autonomy" is used to designate the minimum criteria for an agent's being morally responsible and worthy of moral respect. In another sense, it is used to talk about a certain ideal state, to which we should aspire (Arpaly, 2003, 117-148). I do not mean to be talking about autonomy in the minimal sense, but in the ideal. I grant that putting on a FitBit can be a minimally autonomous act, for which one retains responsibility. I'm interested in showing that value capture damages our status as richly self-governing agents, and block us from the many goods of fully involved, detailed self-governance. As a result, we can be responsible for a undermining our fuller autonomy.7

7 John Christman and Serene Khader have worried that perfectionist theories of relational autonomy lead to the view that oppressed agents don't qualify as autonomous, and so don't qualify for the protections of autonomous agents (Christman, 2004; Khader, 2020). There argument concerns an account of minimal autonomy. I do not claim that value capture breaches minimal autonomy; only that it prevents us from reaching an ideal of autonomy. In this way, the discussion here comes apart from earlier discussions of the relationship between and autonomy and various social intrusions, including deformed desires and subservience (Superson, 2000; Westlund, 2003; Buss, 2005). These discussions largely focus on whether a subservient person, whose subservience has arisen in large part from antagonistic social

Consider, by way of comparison, another discussion about how social conditions can undermine autonomy. According to one line of feminist analysis, decisions made by women under conditions of paternalism can sometimes fail to be genuinely autonomous — like the decision to give up one's career to be a stay-at-home mother.8 The argumentative strategy here is to find a way of discounting those decisions, by showing that they don't issue from the oppressed person's true desires, true values or real will — in other words, that they don't issue from their authentic self. One might claim that such decisions result from desires which have been deformed by circumstances of oppression. As Anita Superson argues, we can tell that deformed desires aren't a person's genuine desires because they emerge from a lack of sufficient self-respect (Superson, 2000). Another discounting strategy is to say that such

---

conditions, can be held responsible for their actions. Anita Superson, for example, is concerned with the question of whether a woman who has been taught her whole life that women should be subservient to their husbands, should be held responsible for her subservience to her husband (Superson, 2010). They have often been less interested in the question about whether such intrusions are compatible with an ideal of autonomy — perhaps because it is clear that they are far from the ideal of autonomy. I think it is obvious that, in most cases, the value captured agent meets the minimal conditions of autonomy and can be held responsible for their actions. The interesting question is about whether value capture blocks or aids the ideal of autonomy.

8 I don't mean to be affirming such accounts. I find criticisms of the view as excessively paternalistic — especially Serene Khader's recent criticisms — to be quite telling (Christman, 2004; Khader, 2020).

decisions are inauthentic because they arise from excess deference. For Westlund, an agent is autonomous when they takes themselves to be in the position of engaging in critical reflection and providing their own reasons. But the deferential person does not — they simply defer, brutely, to another's decisions (Westlund, 2003). Both strategies involve offering a condition for authenticity and autonomy (self-worth, critical reflection), and then showing how agents might fail to meet that condition, and so fails to be autonomous. My strategy is different: I want to show that certain decisions, even if they arise autonomously — from an authentic self — can result in the undermining of future autonomy and authenticity.9 One might have thought that autonomy is intrinsically self-preserving — that if an agent makes a decision autonomously, then that decision cannot diminish the agent's autonomy. I am suggesting, that it is possible to *autonomously damage one's own autonomy* — to make decisions that decrease one's future capacity for full self-governance.

At this point, one could stand firm on some particular theory of ideal autonomy, and insist that value capture is compatible with that ideal. I would like to suggest, instead, that value capture exposes a lacunae in some traditional theories of autonomy. I won't litigate particular theories here — I suspect that a true believer could conjure any number of resources to defend their favored account. But I will make my case by describing, in detail, the loss of self-determination that results from outsourcing one's values. If the reader finds this compelling, then perhaps they will take, as I do, the autonomy loss of value capture as a basic data point, for which we should expect our theories of autonomy to account.

9 Also note that much of the discussion of deformed desire and subservience is a discussion of autonomy in the minimal sense.

To better see the problems involved in outsourcing one's values, let's think about what we stand to gain by keeping the process in-house. When we are the substantive authors of our own values, we can continue to fine-tune our values, and their articulations, in response to our experience of living in the world. Elijah Millgram puts it this way: we cannot determine the right values to have from the top-down — by, say, performing abstract deductions from some a priori conception of the good. Our values should fit our particular personality and psychology, and the environment we find ourselves in. And we find out whether our values are working well for us through the rich feedback of our life experiences — by paying attention to our feelings of boredom, engagement, disinterest, and excitement. We create better-fitting values for ourselves, by adjusting our conception of our values in response to that feedback (Millgram, 1997). Millgram's argument depends on thinking that values can and should change, but we can provide an adaptation for any readers who deny that point. Suppose that there is some set of objectively correct values we should all have — like, health, happiness, knowledge, or what have you. Even if that is true, cognitively limited beings need to have local articulations to guide their pursuit. If one pursued health, one would need to pursue it under some locally articulated set of criteria, be it internal feelings of well-being, or the ability to perform certain fitness challenges. And that heuristic articulation needs to be responsive to local conditions.

So here is a Millgram-inspired argument about *tailoring values*. One's values — or their guiding articulations — help to determine the course one takes in one's life. Different people — with different psychologies, roles, and circumstances — are better suited to different courses of life. We can guide ourselves into a life-course that better suits our own particularities, by adjusting our values — or their articulations — in response to our specific experience of the world. Happiness, satisfaction, and other forms of well-being are complex and phenomenally subtle. The more responsive our values can be to those subtleties, the better our values will support getting at those goods which are entangled with the phenomenally subtle and variable. But externally-sourced values, insofar as the process of their formulation is insensitive to the details of one's particular experience with the world, are thereby less suited to promote a good life in those respects. FitBit's step-count measure leaves out any reference to phenomena like one's aesthetic experience of the landscape, or joy in the grace of your own body. When one has been value captured by FitBit, one is no longer fine-tuning one's values and aims in light of that feedback. Using a FitBit won't help one directly target fitness experiences rich in beauty, grace, or natural wonder, because a FitBit doesn't measure those things — and FitBits doesn't measure them, because they're not the kinds of things that institutional systems and mass-produced objects can easily measure.10

10 Those who are familiar with recent debates over autonomy know that many distinguish between "proceduralist" views of autonomy, where autonomy depends on features of the process by which an agent arrives at their values, etc., and "substantive" views of autonomy, where autonomy depends on whether the agent embraces certain substantive values (Christman, 2004, 148). I have chosen not to discuss substantive views of autonomy, because I think it is relatively easy to show how, under such a view, value capture can lead to a loss of autonomy — that is, when it leads one to fail to value the right things. I myself hold a hybrid view of autonomy, in which procedural and substantive elements are both important. But the argument I offer is pitched towards what I take to be the toughest opponent — that, even by

I've tried here to be as vague as possible about what counts as a good life, so that the reader may plug in their favored account, whether it concerns well-being, happiness, pleasure, or flourishing.[11] My proposal should make sense under any of the standard accounts, so long as that account permits any subtle variations in the way in which different people achieve such goods. Institutional metrics may succeed at targeting the sorts of goods that are relatively invariant between people, but they are not good at guiding us to pursue the goods that are entangled with our particular personality and our specific place in the world.

To sum up: value capture removes our capacity to fine-tune our values in light of our complex and subtle experiences of our particular corner of the world. It thus diminishes our values' capacity to support those goods that are intimately connected to those complex and subtle experiences. This helps us to see where value capture might be more or less harmful. Some goods are relatively universal and invariant — like controlling pandemics or ending

proceduralist standards, value capture leads to a loss of autonomy, insofar as the procedure of deliberative is inattentive to the particular experiences of the deliberator.

11 One thing this argument may not work for is something like an objective-list view of value or well-being — that there are certain things that are simply good for humans, and that we should simply value them. But I think proponents of such a view would have a far easier route to criticizing value capture. It is extremely unlikely that institutional processes would hit the right targets. It is views of human flourishing that depend on the ability for humans to autonomously set their values that are the trickier place from which to mount an assault on value capture.

climate change. In cases of such clear and public goods, the instrumental goods promoted by value capture may outweigh the harms. But other values are more personal — like the values that govern one's aesthetic life, career path, or fitness. Such values aim at goods which are deeply entangled with variable, personal responses. These are precisely where we lose the most, when we buy our values off-the-rack.

I am not suggesting that we all must somehow invent entirely original values for ourselves out of thin air. We are culturally embedded creatures, who are formed in social settings. Rather, I am arguing that, wherever we get the seed of our initial values from, it is good for us to continue to adjust the details of those values over the course of our lives. We should keep a hand on the tiller — listening to how our life is going and fine-tuning the exact formulation of our values in response. But in outsourcing, we give up on that ongoing, participatory tailoring.

The function of metrics

But does value capture always peg us to inflexible, alien interests? Or are some externally-sourced values more intrusive and more rigid than others? To answer that question, I'd like to focus for while on the particular case of value capture by institutional metrics. Metrics, let us say, are quantified representations of value, which emerge to serve the functional interests of large-scale, bureaucratic organization. The problematic features of value capture are sharpest in the cases of metrics. And the intrusiveness of metrics turns out to be intimately related to their basic institutional functions. Thinking about metrics will help us find our way to a more general understanding of value capture.

First, why are quantification representations of value so common? For one thing, technologies like FitBit are part of the intentional effort to use quantified measures to increase motivations. They are, in the modern parlance, *gamifications*. They use the design features of games — like points, leveling up, and achievement badges — to increase motivation for everyday activities, such as in work and education (McGonigal, 2011). Such gamification have a clear potential for value capture. I won't, however, devote much time with such intentional gamifications in the present discussion. Their insidiousness is more obvious, and has been analyzed at length elsewhere (Gabrielle, 2018).12 And it is relatively unmysterious why they are so seductive and appealing. They have been intentionally designed to be so, using an intentionally refined technology of addiction (Schull, 2012).

I will concentrate on the less intentional forms of value capture. Institutions often create quantified, simplified measures of value for institutional purposes other than manipulation and control. Such metrics are often made for administrative efficiency. But, I will suggest, their role in administrative efficiency also makes them particularly seductive and particularly harmful to ingest. One might now protest: perhaps value capture by institutional metrics is only a problem when the metrics are badly formulated or incompetently managed. Couldn't we just build better metrics? But the problem, I think, is not so easy to escape. It is built into the very nature of metrics. The problematic features of metrics — their narrowness, inflexiblility, and hyper-explicitness — are inextricably bound up with their core institutional functions.

Luckily for us, there is a rich body of empirical research on this question, largely

12 <Footnote removed for review>.

conducted in sociology, history, anthropology, and communications. A foundational work here is Theodore Porter's history of quantification culture, *Trust in Numbers*.13 Porter is particularly interested in how quantified forms of justifications, like the cost-benefit analysis, came to dominate politics and management. The goal is not to show that quantification is simply bad. It is to get clear on the relative advantages and disadvantages of qualitative and quantitative forms of knowing, and to figure out when each is most appropriate — and when it is inappropriate. Says Porter: Qualitative ways of knowing are nuanced and context-sensitive. But qualitative information is difficult to manage en masse, and difficult to transfer across contexts. Qualitative evaluations require significant shared background knowledge to adequately interpret. When we transform information from a qualitative to a quantitative format, we strip off much of the nuance, texture, and context-sensitivity. But by doing so, we create a *portable* package of information, which can be easily sent across contexts and understood by people with little shared background. Quantified evaluations can be easily transmitted between people with little shared background, precisely because they have been stripped of context-dependent features. And quantification isolates the more invariant parts of that information, so that the results can be readily aggregated. For this reason, quantitative methods are preferred by large-scale institutions, which must pass information across many levels of hierarchy — between distant administrators with low shared context (Porter, 1995, 3-86). In other words, quantifications are preferred in large-scale institutions precisely because of their narrowness.

13 Though Porter is a historian, he was deeply influenced by Ian Hacking's work in the philosophy of science on the formation of categories and measures.

And quantitative evaluations themselves vary according to their nuance and context-sensitivity. Once, land in England was commonly measured in *hides*. A hide is the amount of land required to support the average family. The hide is a measure which highlights a highly relevant functional quality. The acre is a measure of land size, rather than land function. When a ruler attempts a fair distribution, the measure they use will determine which quality is evenly distributed — in this case, land size versus land functionality. Think about the difference between, say, a king's giving each of his soldiers a hide of land, versus his giving them each ten acres of land. One might think that the hide is a superior measure of functional worth, and so the vastly preferable measure for providing fair compensation. But hides are highly variable in size; determining what counts as a hide requires the application of detailed local knowledge. A hide in a fertile river valley is smaller than a hide in a desert. Hides also vary depending on local weather patterns, game animal migration patterns and more. The hide is a measure that can really only be effectively managed at the administrative periphery — by locals, who know their environment and its inhabitants' typical needs and usage patterns. The hide is impossible to administrate from any sort of distant bureaucracy. So, says Porter, when we shift from small, local, distributed governance to large-scale centralized governance, we inevitably shift from informationally rich — but difficult-to-manage — measures like the hide, to more standardized, but informationally impoverished, measures like the acre.

James Scott calls this the state's view of the world. By 'states', Scott means any large-scale institution, including governments, corporations, and the emergent networked institution of globalized capitalism. States, says Scott, can only manage what they can see, and they can only see that information which has been rendered into a form which can be processed bureaucratically — information that has been standardized and quantified. States can only see those parts of the world which has been rendered *legible* to them (Scott, 1998, 11-83).

Students grades provide a familiar example. In the modern educational environment, student grades are almost always quantified. But we could easily imagine an alternate educational universe in which we didn't offer quantified grades at all. Educators could just offer qualitative evaluation of their students' work — describing the good qualities of the work and its problematic qualities. Such evaluations could easily be tailored to each students' own particular goals. I might give very different suggestions to a nursing student interested in the practical implications for their work than I would to, say, a future lawyer or future journalist. If our goal is simply to educate the student, we don't clearly need any common scale for evaluation. But in our actual world, we must offer a quantified measure of each student's success — a measure which permits us to compare any student with any other. The easy rankability of student efforts is extremely useful to administrators. All of a student's efforts in a class can be expressed in a single number: their grade. And all their individual grades can be aggregated to generate a single number, which represents their entire educational career — a Grade Point Average. And the existence of GPAs is enormously useful for the project of administrating a large-scale educational bureaucracy. They enable all kinds of fast, easy, and objective-seeming manipulations. An admissions officer can arrange the data from every single student applications into a spreadsheet and quickly sort them by GPA. They can create an automatic cut-off point below which student applications are automatically discarded. Sets of student GPAs can be aggregated in order to yield a single number that can be used as a metric of performance for a particular teacher or a school district. Qualitative evaluations of student might be nuanced and context-sensitive — but they are illegible to the large-scale administrative institution.

Finally, these various procedures — data-collection, transformation into standardized inputs, and aggregation — need to be codified into a set of policies that can be readily explicated and transmitted. This is because large-scale institutions depend on being able to train up people from different backgrounds to perform the same sorts of tasks. And their performance needs to be assessable and auditable by others — where those auditors also come from different contexts, and their audit procedures themselves are subject to the same demands of explicability and transmissibility (du Gay, 2000). These various procedures need to be *standardized*. That means that the inputs and processing rules of these procedures need to be regulated across many contexts (Bowker and Star 2000, 13-16).

We can draw, from this mess of observations, some underlying themes. Institutions share a basic functional interest, inherent to the functioning of large-scale administrative systems. They need to manage information across a vast domain. This need arises essentially from the need for an institution to function as a coherent whole. Notice here that I am not presuming that the institution has some interest in controlling or manipulating individuals. Even the most well-intentioned of organizations — like, say, a charitable non-profit — has this same functional interest in information management. The interest arises from the basic conditions of coherent group agency, as instantiated in a policy-based, centralized bureaucracy.

This functional interest is served by two standard mechanisms — quantification and standardization. Institutions need to render the world into a format legible to large-scale institutional information-processing procedures. So institutions need information in quantified and standardized format. Because of their institutional function, these mechanisms — quantification and standardization — tend to have a particular set of features, which make them problematic to internalize as personal values. First, quantified metrics are *narrowed* by design. Only certain things count. Institutional measures need to be to be usable across different contexts. This requires that the measures leave aside contextually variable information, and focus, for their inputs, on context-invariant qualities. As Scott says, the narrowness of the metric creates a narrowness of institutional vision. Institutions can only see, process, and act on parts of the world that are counted by their metrics. Anything that doesn't impinge on those metrics is invisible at an institutional level. In value capture, we internalize those narrowed metrics, thus narrowing our values. And, insofar as our values drive our attention, then the value captured will be subject to an analogous effect to Scott's narrowed institutional vision. It's not that we literally don't see things that fall outside our narrowed values, but we won't devote much energy to them, or dismiss them as unimportant.14 Think here of the businessperson who thinks that only money matters, and who immediately dismisses from mind any unprofitable ventures — like art or philosophy.

Next, such institutional metrics present values in highly explicated, *finished* form. They are typically formulated so as to resist re-interpretation. Pre-institutionalized values are typically expressed in an open-ended manner. A concept like "health" or "fitness" or "a good education" admits of different interpretations. Different people may work out their own interpretation of what counts as a good education — and so evaluate their understanding of the term. You want to know more useful things, I want to indulge my sense of curiosity — both are viable understandings of how one might go about getting an education. But step-

14 <Footnote removed for review.>

counts and law school rankings do not admit of such variability. The mode of assessment is rigid. In fact, says Porter, that the process of quantification is useful to large-scale institutions in part because because it reformulates information so as to remove the need for interpretation (Porter, 1995, 21-29). Standardization is required for informational portability — and standardization requires rigidity.

And those off-the-rack values usually come *embedded* in institutional infrastructures, institutional language, and mass technologies, so as to resist further tailoring. We don't have the power to fine-tune the innards of such institutional values. We can't tinker with the way Twitter counts Likes, nor adjust with the *USN&WR*'s ranking algorithm. They are hard-wired into external systems. This rigidity and uncustomizability of measures and metrics is no accident; it is essential to their institutional function. Standardization enables easy communication and ready aggregation — but, to do so it must resist individual customization.

To summarize: institutional metrics are designed according to *alien interests*.15 They have, in fact, been fine-tuned and adjusted, but according to interests that are not our own. I

---

15 (Owen and Cribb, 2019) make a similar point in their analysis of FitBit technologies (32-35). They distinguish, however, between procedural autonomy — which involves internal deliberative processes — and substantive autonomy, which involves one's ability to actually acton and bring to fruition one's decisions. They say that, by and large, self-tracking technologies like FitBit may aid procedural autonomy, but cannot aid substantive autonomy, since such technologies can't fix large-scale social inequities. My argument is that such technologies also significantly undermine procedural autonomy.

am not presupposing, here, that institutions must have malevolent intent — like an interest in domination, control, or power. But institutional metrics are typically formulated to fit the basic functional needs of large-scale institutions — for easy recording in institutional memory, for transmission across bureaucratic layers, and for manipulability by institutional methods. When we internalize institutional values, we are letting such interests play a powerful role in the formulation of our own values. Value capture gets us to take an institution's eye view on ourselves — to evaluate ourselves and our activities in institutional terms.

We have much to gain by fine-tuning our values, fitting them with our psychology and world. Institutional metrics are tuned, not to an individual's rich and particular experience of the world, but to the needs of information processing at a mass scale. In value capture by institutional metrics, our values become rigidly tied to an external expression — and that rigidity arises, in significant part, from the institutions's interest in large-scale informational management. Their form — and their embeddedness in external systems — resists attempts at digestion and customization by the agent. Their alienness resists domestication.16

---

16 One might ask what relationship this view has with various forms of alienation critique. Though my analysis here is obviously similar, in spirit, to the general themes of alienation critique, I avoid use of the term "alienation" because my analysis here differs, in key respects and in many details, traditional alienation critiques. As Rahel Jaeggi says, many forms of alienation critique involve views where the alienated agent as divided against themselves, as unable to identify with their work, as diffident and depressed (Jaeggi, 2014). But the value captured agent can be wholehearted (think of the capitalist all-in for money), fully identified

The seductiveness of metrics

Of course, we often have to use such metrics when we work within, and next to, institutions. But we could do so while keeping them at emotional arm's length. We could employ them in our reports and our requests for funding, but only as the trade language of bureaucracies. Why might we ever take the further step and internalize them? The answer comes in several stages.

First, quantifications, in and of themselves, are cognitively seductive in their clarity and crispness. Many people seem to trust quantified data simply because it is quantified. And we should trust data, when it has been generated using reliable methods. However, the mere quantified format itself often seems to generate trust, regardless of quality of the underlying methodology. But obviously, mere presentation in a quantified format does not offer any guarantee of reliability. So, insofar as we trust from the bare fact of quantified presentation, then that trust is unwarranted. And Porter, Merry, and Espeland and Sauder provide

---

with their work, energized and motivated. They are not divided against themselves; rather, they are simplified, where that simplification has been guided along institutional lines. Notice, furthermore, the difference between my analysis and the traditional Marxist alienation critique. It is possible to be value captured by a fully socialist bureaucracy. Here, I am aligned more with Scott's particular version of neo-Foucaltian critique, then with Marx. For Scott. both globalized capitalism and centralized communism share an interest in rendering the world legible into the terms which they can process and act upon.

evidence aplenty that bare fact of quantification actually does, in fact, generate such unwarranted credibility. To put in the contemporary parlance, the excessive credibility given to quantified data counts as a form of epistemic injustice or epistemic oppression (Fricker, 2007; Dotson, 2014). It harms those who are unwilling or unable to present their information in such quantified form, preventing them from being appropriately recognized as sources of information. And insofar as quantified data tends to emerge from certain sorts of institutions, then those institutions themselves are the beneficiaries of epistemic injustice.

Why might the mere presentation of information in quantified form invite such excess credibility? One familiar suggestion is that numbers carry with them, through their association with the sciences, an aura of authority. I'd like to suggest another mechanism, arising from *cognitive fluency* — a phenomenon which has been well-documented by cognitive psychologists. Cognitive fluency is the "subjective experience of ease or difficulty with which we are able to process information" (Oppenheimer 2008, 237). As it turns out, we often use cognitive fluency as an epistemic heuristic. The easier an idea is for us to comprehend, the more likely we are to accept it as true. This is sometimes a useful heuristic. We are typically better at processing information in domains where we have expertise, so ease-of-comprehension is somewhat correlated with correctness. But the heuristic is far from perfect, as cognitive psychologists have amply demonstrated. First, we seem more willing to accept an idea simply because it is familiar. Second, we are more likely to accept claims presented in a more legible font. But, obviously, the bare fact of repetition or graphic legibility has no direct bearing on truth. In both cases, using a cognitive fluency heuristic results in a mistaken degree of trust (Reber and Unkelbach, 2010).17

We should, then, experience a cognitive fluency effect with anything with which are familiar. And we are extremely familiar with numbers. They are the universal abstraction. Information presented in quantified form thus wears an extremely familiar face. So, the fluency heuristic can lead us astray with quantification, just as it does with fonts. This offers an explanation for the unwarranted credibility of quantified values. Fluency may bring somebody to accept an quantified evaluation of value over a more inchoate one — like accepting the *USN&WR*'s clear presentation of a ranking, over one's own internal sense of, say, fit with a law school's culture. And insofar as the quantified presentation is more likely to emerge from external and institutional sources, then the fluency effect gives an unwarranted credibility boost to such sources.

But it's not just that metrics are quantified — it is that they are *standardized.* Once our values are standardized, then we can communicate our justifications with extraordinary ease. This engineered communicability grants a further credibility advantage to claims made in the evaluative language of those metrics. After all, our ability to make ourselves understood to others can be a sign that our own understanding is good.18 And metrics are, by their very nature, easier to understand across contexts. But there is a gap between

17 <footnote redacted for review>.

18 I am relying here on the literature from the philosophy of science's investigation of understanding. According to the standard account, one of the signs of understanding is the ability to communicate that understanding to others (Strevens, 2013). <Sentence removed for review>.

communicability and epistemic worth, and that gap can be exploited. As Porter makes clear, institutional metrics trade away informational nuance, richness, and contextual sensitivity for the sake of easy portability across communicative contexts. Metrics, then, in virtue of their basic institutional function, also function to precisely exploit the gap between communicability and epistemic worth. When we standardize metrics, we engineer communicability and familiarity — which invites excess trust.

Let's turn now from the discussion of trust and fluency to an very different sort of mechanism. Quantifications can be seductive because they offer us the pleasures of *value clarity*. When we internalize them, our value landscape becomes simpler and easier to navigate. We are tempted to take them on, because they offer us hedonistic rewards, in exchange for simplifying our values along certain lines.

This line of argument draws on C. Thi Nguyen's account of the motivational structure of games (Nguyen 2019, 2020). In games, says Nguyen, we take on artificially constructed goals. In ordinary life, our goals and values are often complex and subtle. It is often hard to explain our values clearly, hard to adjudicate conflicts between values, and hard to figure out if we've actually achieved what we value. But in games, values are easy. They are clearly articulated, with crisp criteria for application. In games, we know exactly what we should be doing, and exactly how well we've done. Games offer us a momentary refuge from the nauseating complexity of real world values. They are an existential balm.

Nguyen's account thus offers us a second mechanism for the seductiveness of quantification. We can gain a hedonic reward for internalizing simplified values. When we come to value a simplified goal in a non-game activity, we bring the pleasures of value clarity into the real world. Our purposes become clearer, our degree of success becomes more obvious, and our achievements become more readily rankable. But to get those pleasures, we need to simplify the target. And this helps to explain why it can be so tempting to internalize institutional metrics. Metrics are narrowed and finished. When we internalize such clarified metrics, we can eliminate our struggles with the ambiguity and complexity of our values. Metrics may not have been explicitly made for gamification, but the institutional pressures on the generation of metrics make them function as pleasingly game-like goals.

And value clarity effect becomes even more powerful when that clarity is standardized. After all, the existential burden of our complex values is not merely a personal affair. The buzzing value plurality of the social world can also be profoundly painful. Even if we have a good grasp on our own values, explaining ourselves to others can be a grueling experience. There is, so often, a vast gap between our differing experiences of value. Try explaining to another person your profound love of some weird old comedy, or why a sour cabbage casserole makes you feel so comforted on the bleakest of days. Try explaining why a particularly acid passage of Elizabeth Anscombe's fills you with such glee, or why you never quite got along with running, but rock climbing makes you feel so amazing. Sometimes we can make ourselves understood; but so often, the gap between us is too great. So much of our sense of value arises from our particular experiences, the long life we've lead, our twisty paths to self-understanding and world-loving — that explaining the whole mess others is often beyond our capacities.

But institutionalized values offer us an experience of social value clarity. If an institution offers us a pre-fabricated metric of value, and we collectively internalize it, then we will make easy sense to each other. Perhaps that pre-fabricated metric is citation rates, or Twitter followers, or GPA, or how good of a law school you got into. In any case, once we internalize that value together, much of the existential friction of social life suddenly disappears. Metrics create a common currency for justification. I no longer need to struggle to explain my way of valuing to others, or to understand their way of caring about the world. Justification becomes easy, because metrics offer a pre-engineered system of aligned value. Metrics offer, not just a personal form of value clarity, but *social value clarity*.

Let's take a step back. Is it some wild accident that institutional metrics turn out to be so seductive? I suspect not — though I can only offer a brief sketch here. Rational agents often need clearly articulated policies to function — including policies about what our goals are, and how we are to evaluate our progress towards those goals. Clearly articulated policies ensure reasonably fast decision making that is consistent over time. As Michael Bratman argues, such policies play an integral part in our being able to maintain coherent agency over time (Bratman, 1987; Holton, 2009; Andreou, 2010). Policies are desirable for large-scale institutions for similar reasons, since institutions also need to ensure relatively quick and consistent decision making across a large and scattered structure, to enable cohesive collective action.19

But the nature of large institutions requires that we heighten the clarity and explicitness of those policies, in order for them to function across consistently across the whole. The policies I set for myself can hinge on my own peculiar sensitivities and ways of understanding the world. A coherent policy for me is: "Exercise every day until I start to get that pleasant warm cheerful feeling," which works because I can consistently recognize that

19 This comment relies on the extensive recent literature on group agency, including (List and Pettit, 2011; Gilbert, 2013; Rovane, 2019)

pleasant warm feeling. (Another coherent policy for me is, "Cocktails before 6 PM only when I really, really, really need it.") But such policies won't work for large-scale institutions, because criteria like "a pleasant warm cheerful feeling" cannot be written into institutional policy, nor could they be reliably applied by different people across the institution. Institutional policies need to be hyper-explicated so that they may be executed by a wide variety of people, hired from a variety of backgrounds. They need to be, to adapt Porter's language, procedures that are *portable* between many contexts. So the features of policies that help us to be stable and coherent individuals are exaggerated in institutional policies, and so they can be appealing to internalize. It is even easier to act clearly and consistently when we adopt such a hyper-explicit policy. However, in adopting them, we are giving up on the kinds of policies that hinge on sensitivity to subtle internal phenomenon.20

When is it reasonable to outsource?

One might now reasonably worry that I am placing an impossibly high standard on individual autonomy. After all, an ordinary life is full of perfectly reasonable — and perfectly humdrum — value outsourcing. For one thing, we often trust others to find our way into new forms of value. Once upon a time, I didn't get jazz. But I trusted a friend to show me how to appreciate it. I listened to him talk about jazz and I tried to follow him in, trying to find my

20 I owe this way of connecting institutional policies and individual planning agencies to a suggestion of <redacted>. For a further discussion of problems with the explicit policies in group agents, see (Nguyen, 2019).

way to seeing the beauty that he saw.21

But notice that this kind of trusting relationship is very different from outsourcing. In trusting my friend, I may take certain actions purely because he told me too. I devote some attention to John Coltrane's *Giant Steps* just because my friend thinks it's so amazing. But, hopefully, my trust in him is merely a starting point. I doubt either of us would be very happy if, ten years down the line, I continued to imitate his every judgment — if my love for jazz still mirrored his in every detail. The goal, I take it, is for me to use his sense of value of a springboard for me to help me find my own way. Here, I am not outsourcing, but learning.

But surely we engage in real value outsourcing all the time, because we simply don't have the time to think through everything for ourselves. When I bought my dishwasher, I just looked at some reviews and bought what the experts recommended. I am outsourcing my determination of "best" dishwasher to those experts, it seems, just like the prospective law school student outsources their determination of the "best" law school to the *USN&WR*.

But outsourcing my dishwasher values seems significantly different from outsourcing my values in health, education, or career satisfaction. First, dishwasher values are fairly thin. The functionality of dishwashers is, if not utterly one-dimensional, far less multidimensional than the value of an education, health, or career satisfaction. Dishwashers are simple tools with clear, generally agreed-upon functions. The good of that functionality has low entanglement

21 See Tal Brewer's rich discussion of how we slowly uncover the value of an activity over the course of doing it (Brewer, 2011). I am adding here only the suggestion that often we trust others to help us find our way in. (Agnes Callard has also provided a recent commentary on this topic (Callard, 2018).)

with the complex, subtle, and variable phenomena of our individual experience. Such objects —which aim at simple, impersonal targets — are good candidates for outsourcing, when we need to save on some cognitive resources.

And these cases seem unproblematic because they concern matters at the outer periphery of our values. To outsource a value is to put stop deliberating about it — to withdraw one's attentions from it. And we cognitively limited beings do need to withdraw from certain domains. But the best place to do so is about those things we care about the least. We want to spend our cognitive resources where it counts. And when we are invested in something, and deeply involved with it, we have far more to gain from fine-tuning our values in that space — precisely because those domains loom large in our experience and our well-being. When the person whose life is devoted towards a legal career lets the *USN&WR* set their notion of the best legal education; when the dedicated scholar lets citation rates determine the content and direction of their research — this is where the lack of a personal fine-tuning will really matter.

Reflective control

Let's consider one last, and very important, objection. Why think that value capture is any kind of trap at all? That is: why not think that I am simply aligning myself with certain external values for the moment, which I can, at any moment, pivot away from? After all, cognitively limited beings like us often find it quite difficult to always live in the full light of our complex values. So we must conduct much of their day-to-day thinking through proxies and heuristics. Step-counts aren't the same as health, but "health" is a fuzzy target, and hard to aim at in the daily grind of ordinary decision-making. So, better to aim at step-counts in daily life. I have not lost my autonomy so long as I retain reflective control over those proxies.

This is all entirely reasonable. A fully autonomous being could certainly use FitBit's step-counts as a local proxy for health. But to remain autonomous here would require proper management. Proxies can help us achieve richer values, but, of course, a proxy and the value which it stands in for can come apart. To be fully autonomous, they would need to occasionally reflect on whether pursuing the proxy value actually helps them get what they really value. For example: so long as pursuing more step-counts lead, in the long run, to feeling healthier and being fitter, then it's a decent proxy. But suppose my knee started to hurt, and my constant attempts to up my step-count just caused more and more knee damage, until I was in constant pain. (If the knee example seems far-fetched, think about the sorrow and life-agonies that have been induced by people in pursuit of higher citation rates, more Twitter followers, or more highly ranked law schools.) This is a failure of autonomy because it is a failure to control my proxies, in light of my full and larger values. I may have chosen the proxy initially, but I am failing to adapt my use of proxies responsively. I am, we might say, self-determined, but not self-determining.

And it's more than simply checking my proxy in light of some larger and stable value. In many cases, when I try out and reflect on various proxies, I'm actually exploring an activity to find what values it might offer me. Most of our activities can offer different values for different people and different approaches. Suppose I try to discover a way of exercise that I can really love and inhabit. I might first aim to up my step-count, and, after a brief burst of motivation, eventually find myself miserable and bored, pounding out aimless miles just to make the numbers. I might change to a different proxy target, like trying to up my speed and win some races. This might provide me more difficulty and intensity — but it also might pall. I might switch to trail-running and aim to try out as many different trails as possible, to explore a variety of terrains in a variety of seasons. This makes tracking mileage more complicated and uneven — since the surfaces are so variable — but also gives me more of a sense of aesthetic exploration. Or I might ditch the whole running thing and take up yoga or pole-dancing. In each case, a certain degree of openness towards what my larger values are — a spirit of discovery — lets me adjust my proxies, trying out life under them for a while, getting feedback, and so exploring different ways I can find value in an activity.22

So the real question is: why aren't value capture cases just cases of well-managed proxies? Here's a cheap answer: well-managed proxies aren't value capture by definition. Value capture occurs when an externally-sourced value dominates practical reasoning, and the cases of well-managed proxies are not such cases. In the well-managed proxy cases, my lasting values — and my sense of deeper value — dominate the proxies. And many of the cases we've seen aren't marked by such reflective control.

The cheap answer is unsatisfying, but it exposes the real question. How is value capture even possible? In my portrait, value capture undermines autonomy because it leads to a particular kind of loss of control. One might worry that ceding control in this way is impossible for a rational agent — that a rational agent is always constituted such as to be able to take back control over their values and their value articulations. If I live by it, haven't I chosen it?

22 I am strongly influenced here by Brewer's account, and by Millgram's discussion of developing values through practical experience (Millgram 1997, 2004).

But good proxy-management depends on stepping back from the proxy and surveying it from the standpoint of one's larger and richer values. If one avoids that reflective step, then one is no longer managing the proxy. One could be making a decision for which one is fully responsible — to commit to a proxy in a way which avoids future reflection on one's values — but the effect of that decision is to give up further responsive control over that value. And it is not like, once one has ceded control, control can come back in a moment. Reflectiveness, after all, is not just a brute disposition. It is a practice, a skill, a cultivated habit of mind. And you can get out of the habit of reflection. You can develop, in its stead, the habit of ingesting your values from the outside, of uncarefully snapping them up from the systems and technologies which surround you.23 When you have gotten into the habit of acquiring one's values readymade and pre-fabricated, you can get to used to the ease — and get out of practice with the painful struggle of doing it for yourself.

But, one might say, the person that uses a FitBit does get up every morning and decide to put their FitBit back on their wrist.24 How could using FitBit as a proxy for one's values every be anything other than autonomous, since one consents to participate in it on a regular basis? This way of putting the point, however, depends on a thin and binary notion of autonomy. The person that decides to put their FitBit in the morning may, indeed, be deciding to use FitBit as their articulation of fitness values for the day. But their relationship to that value is interestingly distanced. They have said yes to a whole package, but they have not authored the particular details of that package. They have the minimal autonomy of having said yes,

23 <redacted for review>

24 I owe this lovely way of putting the point to <redacted>.

but not the more substantive autonomy of having decided, on the details of that package. They lack *granular control* over the details of the package.

Their relationship to their values might be compared to one's relationship to an End-User License Agreement. When we sign a EULA, we are restricted to a binary decision about accepting or rejecting the whole package, with no capacity for fine-tuned authorship or adjustment. Such a person can be said to have consented in the minimal sense. After all, when you sign a EULA, you can be held responsible for that fact that you did, in fact, sign it. But you're lacking in a more richly inhabited form of autonomy over your values, where you have fine-grained control over the details of the package. This more substantive conception of autonomy — where we participate in working out, investigating, adjusting, and even authoring the particular details of our values — is not a binary notion. It is an ideal, and one which we, as limited beings, can never wholly realize. But to permit value capture is to retreat to the thinnest possible form of autonomy — the minimal autonomy, where what one can say is that one has, at least, smiled and nodded and signed on the dotted line. Value capture puts you into the same relationship to the particular contents of your values as you do towards the particular contents of your smartphone's EULA.

Towards value federalism

So far, my discussion has focused on cases in which an individual becomes value captured by values presented by a larger group. But is it also possible for groups to be value captured by other groups?25 At first glance, it seems entirely possible. Groups can have values.26 Small groups can tailor their values to their own particular needs, which can be swamped by the larger presentations of larger or more prominent entities. A particular philosophy department might might be value-captured by its university's presentation of particular metrics for student success. A university might be value captured by a magazine's ranking of schools. A corporation start out dedicated to providing an elegant and beautiful user experience, and become value captured by its shareholder value. A non-profit might start out dedicated to changing the world for the better, but come to be value captured by its status on some list of most effective charities.

But the mechanisms and possible harms such group-level value capture must be markedly different from those of the individual cases which we have so far discussed. For one thing, group-level value capture doesn't involve the essential transition between internal phenomena and external policy that individual-level value capture does. When an individual is value captured by the metrics created by a large institution, their values are no longer responsive to the special inner phenomena accessible by an individual. But the loss of such sensitivity to the inner can't be what's wrong with group-level value capture. So why might such group-level value capture be harmful? A full answer will depend on the resolution of some thorny issues about the nature of shared values, collective intentions, and group

25 I owe this question, and other insights that have informed this section, to <redacted>.

26 For a minimal account of an entity that might count as a group value, see any account of shared intention (Rovane, 1997; List and Pettit, 2011; Gilbert, 2013; Bratman, 2014). I find particularly useful the account of shared valuing in (Hedahl and Huebner, 2018).

agency. Here I will give a preliminary sketch, which will point the way towards a fuller account of value capture and its harms.

Let's try to adapt the observations about individual tailoring and value capture to the case of group values. The problem is easiest to see when we think about value capture between vastly different scales of groups. My spouse and I form a very small group agent, with some shared values. Before we had children, we actually spent a lot of time talking about and settling on some our shared goals and values. But once the kids showed up, we adjusted those goals, and the underlying conception of what we valued about our family. We did it in response to the density of new and unexpected experience, the thousand new emotions and responses we had to life with children. Notice, here, that though the value is shared, it is still substantially responsive to each of the individual's own concerns. Our process of setting our shared values needs to be responsive to both of us, so there will be some compromise. But the values can still it can be substantially tailored to our various needs. However, if we were value-captured by some external metric for success in a family — say, for example, by the metric in terms of our children's school status or eventual income — such value capture would bring in a far less tailored value. Those metrics represent external interests, and are far less unresponsive to our particular psychologies than the small-scale, intimate shared values we cooked up together.

What if I get to participate in the setting of those metrics? Here, scale matters. Suppose I am a member of some very large group. Even if I participate in the process of setting those values, so long as the group is sufficiently large number of people, then the values it declares will be largely insensitive to my particular needs. One might think that each United States citizen participates in the setting of the values of the United States through voting and participation in public discourse. But I don't take it that what comes out the other end is particularly responsive to any individual's particular psychology.

We can generalize the story, so that we don't need to make reference to individuals at all. Different groups have values. Those values can be tailored to their particular local conditions. The well-adapted values of, say, an musical community working in an environment with only memory and oral transmission will turn out to be quite different from the values of a musical community which uses written scores, and different again from one working with simple recording technology, and different again when working with the ability to remix. The value of improvisation, as we find it in jazz, is well-attuned to the existence of recording technology. The ability to sight-read a score is far less valued in a community working with sampling and remixing technologies. But the values that arise in a larger-scale institution — like, say, an international music recording conglomerate — are far less responsive to those small-scale, local considerations. So value capture of small-scale communities by the expressed values of larger-scale communities can introduce a similar form of unresponsiveness to local conditions. Similarly, the various departments at different universities of a university might have their own values, finely tuned to their attempts to educate a particular student body to their own particular discipline — but those fine tunings can be swamped by university-wide, or system-wide metrics of student satisfaction and student success.

This suggests a complex picture about how values should be fine-tuned. It is not that there is some special magic to individual values. Rather, there are different kinds of groups, of varying sizes, each with differing needs. Some kinds of values are perhaps best fine-tuned to the individual — those that are most entangled with individual phenomena, like values involving art, romantic love, and fitness. Some kinds of values are best shared, but at smaller scales — like the shared values of a family, neighborhood, local musical scene, or city. But some values are best instantiated when shared across enormous groups, like multinational corporations or nations, or the collective community of scientists Each scale of value has its own trade-offs. The larger-scaled values are good for those efforts that are best pursued through large-scale harmonized, collective efforts. But those larger-scale values are incapable of precise tailoring to the details of local conditions.

What emerges is a picture of what we might call *value federalism*. Just as there is a reason to have different governments at different scales — international, national, state, and city — there is also a reason to have values at different scales. What massively-scaled values gain in their capacity to organize great collective effort, they lose in their capacity to tailor to individual lives and smaller-scaled local communal efforts. So we, as individuals, have reason to participate in values at various levels of scale.27 Value capture truly hurts us when we let larger-scale values wholly swamp our individual, and smaller-scaled values.

Legible values

---

27 One way for an individual to participate in a value is to actually take it on; another way is to use it as a guide for acting within a role, while keeping at at arm's length from one's individual values. Consider, for the latter option, Carol Rovane's account of how individuals can "swap in" other agencies when acting in particular roles, as an individual might swap in the roles of their group when acting as an officer of that group <cite>.

Perhaps, one might claim, value capture is actually agency at its best. David Velleman, for example, suggests that that the constitutive aim of agency is understandability. In his earlier work, he argues that the constitutive aim of agency is self-comprehension — that we regulate our conception of our values and our actions to fit each other, so that we can make sense to ourselves (Velleman 1989, 173-186; 2000, 193-199; 2006, 204). In more recent writing, he adds a social dimension. The constitutive aim of self-comprehension, says Velleman, generates a second aim of being comprehensible to others. We want our pictures of our values and motivations to match the picture that others have of us. This is to avoid double-bookkeeping — the exhausting effort of keeping in mind my picture of my values, and your picture of my values, and all the complexities that compound from that. We want to make our value pictures to cohere, to reduce our cognitive load (Velleman, 2009, 64-68). And if our values come in the same terms — if we value the same things, expressed in the same ways, and evaluated in the same ways, then our cognitive load will be ever-more reduced. We will be readily comprehensible to ourselves and to others, in efficiently consistent terms.

Value capture makes self-understanding easier. If Velleman's account of agency and autonomy is right, then value capture is not only permissible, but ideal. A self with complex and subtle interests is one that is hard to understand. A self that has been re-engineered to value clear and simple metrics is a self that is easy to understand. And it is a self that is optimized for interpersonal coherence. When my interest in hiking proceeds from a strange bundle of inarticulate cravings, satisfactions, and subtle aesthetic pleasures, then self-understanding will be hard-fought and tenuous — and explaining myself will be even harder. But if my interest in exercise can be expressed in terms of step-counts, then it will be easy for me to choose actions which can be readily understood and explained.

But I find myself pulled in the opposite direction. Value capture, I think, exposes the deep problem with Velleman's account — that his conception is of an agency unmoored. For Velleman, there is no recalcitrant self that is to be understood or which demands loyalty. There is only a manipulable package of self-conceptions and intentions and actions that need to be brought into coordination.28 So there is nothing stopping me from re-engineering my values to make them easier-to-apply. In other words, nothing is keeping me from *gaming* the activity of self-understanding by simplifying my self-conception. Of course, the Vellemanian agent's self-understanding must match the complex phenomenology of its experience — its felt joys and confusions, its unexpected pleasures. But nothing is stopping the Vellemanian agent from trying, within the bounds of its own psychology, to re-interpret those joys and confusions in the direction of easy comprehensibility, and trying to simplify its phenomenology in the pursuit of self-understanding. If the goal of self-understanding is not to get at some kind of authentic self, but simply to maximize the fit between self-conception and self, then an excellent strategy would be to radically simplify both self and self-conception. And if we have the further goal of maximizing the fit between self-conception and others' conception of us, then we should point that simplification on any clear external signal. Institutional metrics are technologies of value-alignment, and the Vellemanian agent

---

28 Westlund offers a similar criticism of Frankfurt and Bratman-style theories, which make the ideal agent the well-integrated one: "A self—abnegator who finds herself deeply ambivalent about her deference might well strike us as being *closer* to achieving self-government in choice and action than her well-integrated and internally consistent counterpart…" (Westlund, 2003, 491).

has every reason to make full use of those technologies.

In Millgram's account, values should be adjusted in response to the details of our own experience. The more sensitive and responsive we are to the details of our own experience, the better our values will fit our local conditions. The Vellemanian agent, on the other hand, has reason to desensitize themselves to their experiences, to remake themselves according to a pre-made and public standard and forget what falls outside. If one wishes to resist such re-engineering, it will be out of a loyalty to some sort of more mysterious and inchoate experience — something that might resist facile expression and ready explication. And it will be out of a sense that one's values should be developed through the continuous monitoring and awareness of and responsiveness to that inchoate self, and its relationship with the world.

If one were drawn to the account of value capture that I've offered — and shared the sense that value capture lessens autonomy — then one might also be tempted to give the following diagnosis of the appeal of the Vellemanian story. Let's return to Scott's discussion of the state's interests. In Scott's account, states can only see and manage the parts of the world which are *legible* to them. They need information parsed into their preferred formats, so it can be processed through a large-scale bureaucratic structure of information processing and decision-making. They cannot process sensitive, subtle, context-sensitive information — so they cannot see into spaces characterized by such delicate qualities. Thus, says Scott, institutions have an interest in remaking the world into a more legible form. Scott offers a host of examples: German scientific forestry and its interest in decreasing biodiversity and rendering the forest into neat, regular rows. The ironing out of complex, locally-derived, multi-layered land-use systems into simple laws of ownership. The re-ordering of cities from dense labyrinths into neat grids. The logic of bureaucracy and the logic of capitalism converge, for Scott, on a common desideratum: that the world be made measurable in terms that might be readily absorbed by large, organized information-processing systems.

We can extend Scott's account to offer a diagnosis of Velleman's picture of agency. According to Velleman, the constitutive goal of agency of self-understanding, which generates the further goal of making the self understandable to others. Notice that these goals are quite consonant with the state's interest in legibility. From the state's perspective, people with subtle, variable, complex and inchoate interests are hard to track, predict, and manage. They are illegible. But if we have all converged on an easily publicizable, quantified format for our values, then our values and interests will, themselves, be more legible to the state. The state would vastly prefer that we be Vellemanian agents, motivated to bring our values into its sight. Our interest in convergence, then, can be nicely explained as an internalization of the state's interests. In other words, our very interest in being so readily understood may itself arise from a kind of value capture.

This suggests another, larger view of the function of value capture and autonomy reduction — of why those tendencies might serve institutional interests. One of the highlights of Scott's *Seeing Like a State* is a detailed and gripping story about the growth of large-scale agriculture and its entanglement with state-level modes of oversight. Large-scale agriculture, says Scott, favors low-diversity monocropping — fields of wheat, neatly arranged — rather than complex layered polycropping of many diverse species, grown together so that their various functions interlock. Polycropping is actually more productive. Tall crops shade short crops, one plant combats erosion and another restores nitrogen to the soil. But polycropping's success depends on fine-tuning the arrangement of diverse crops to the local soil, weather, and the particular culture of the farmers. Polycropping tend to need lots of local autonomous decisions on the parts of farmers. But those decisions are non-standardized and hard to aggregate — and thus unmanagable from the state's-eye view. Low-diversity monocropping, on the other hand, is manageable from the top-down (270-282).

A central theme of Scott's discussion is that the enforcement of schemes of legibility on the world tends to drive skill and agency from the local, contextual periphery towards the organizational center. When land-holding rights are complex, nuanced, and highly localized, then the skill to navigate them — and the power that goes with those skills — is largely held by local farmers. When those land-holding rights become evened out and organized according to centralized policies, expertise transferred from the local farmers to a small cadre of centralized bureaucrats and lawyers (72-83). The process of rendering the world legible — standardized, quantified, and manageable — tends to de-skill most workers and render them fungible (250-252). The imposition of large-scale monocropping on what was before a diverse, variable, and localized set of farming practices ends up changing the power dynamics. Before centralized monocropping, each farmer was a particular expert in their own land, making any number of complex decisions in response to their particular environment. After monocropping, individual farmers are supposed to execute policies as consistently as possible, in environments that are re-shaped to be as similar as possible. This makes the individuals essentially interchangeable, and renders their local knowledge useless. The important skills and powers are transferred, instead, to a smaller cadre of distant agricultural scientists and policy-makers. One can see this as a kind of transfer of autonomy — where the legiblization of the world drives autonomy from the distributed periphery, concentrating it at the administrative center.

Large-scale agriculture also prefers legible targets. It prefers to target simple-to-measure qualities, like total yield, over subtle and hard-to-measure qualities, like freshness or a particular delicate aroma (289). So our large-scale agri-business would prefer that its customers didn't care about those difficult-to-target qualities, and valued instead those easy those qualities that are easy to target on a mass scale. Even worse: imagine a world in which individuals autonomously developed their own tastes, exploring different tangents and searching for the subtle qualities that best suited them. It would be very difficult for a large-scale institution to cater to people's tastes in such a world. Better, instead, if our tastes were regularized, standardized, and predictable, so that the institutions could reliably satisfy us. Institutions should want legibility not only of their own ability to manufacture products, but also of the patterns of consumption over those products. So institutions should want our desires and interests to be legible, too. They should want our values to be expressed in terms that can be understood at an institutional level. My suggestion, then, is this addition to Scott: the institutional drive for legibility extends, not just to the material world, but to people's values.

If you buy Scott's story, then the existence of value capture becomes entirely unsurprising. So long as institutions have an interest in managing the world, then they have an interest in managing our own values and interests. They have an interest in using whatever tools they can — gamification, seductive clarity, technology — to get individual values to take forms that are analyzable to the institution. They would be particularly interested in ensuring that individual values aligned with taxonomies usable at the institutional scale. Obviously that interest is opposed to the autonomous development of personalized and peculiar values. The drive to legibility, in all its forms, is one which has an interest in driving autonomy out of the distributed periphery — in centralizing control, and standardizing the elements of the world. The last frontier of legibility is the human soul.

Bibliography

Andreou, C. (2010). Coping with Procrastination. In C. Andreou & M. D. White (Eds.), *The Thief of Time*.

Arpaly, N. (2004). *Unprincipled Virtue: An Inquiry Into Moral Agency* (1st edition). Oxford University Press.

Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences* (Revised edition). The MIT Press.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Harvard University Press.

Bratman, M. E. (2014). *Shared Agency: A Planning Theory Of Acting Together* (1st edition). Oxford University Press.

Brewer, T. (2011). *The Retrieval of Ethics* (1 edition). Oxford University Press.

Buss, S. (2005). Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints. *Ethics*, *115*(2), 195–235. https://doi.org/10.1086/426304

Callard, A. (2018). *Aspiration: The Agency of Becoming*. Oxford University Press.

Christman, J. (2004). Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. *Philosophical Studies*, *117*(1–2), 143–164. https://doi.org/10.1023/B:PHIL.0000014532.56866.5c

Dotson, K. (2014). Conceptualizing Epistemic Oppression. *Social Epistemology*, *28*(2), 115–138. https://doi.org/10.1080/02691728.2013.782585

du Gay, P. (2000). *In Praise of Bureaucracy: Weber - Organization - Ethics*. SAGE.

Elster, J. (1998). *Ulysses and the Sirens: Studies In Rationality And Irrationality* (Revised, Subsequent Edition). Cambridge University Press.

Espeland, W. N., & Sauder, M. (2016). *Engines of Anxiety: Academic Rankings, Reputation, and Accountability* (1 edition). Russell Sage Foundation.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Clarendon Press.

Gabrielle, V. (2018, October 10). How employers have gamified work for maximum profit. *Aeon Magazine*. https://aeon.co/essays/how-employers-have-gamified-work-for-maximum-profit

Gilbert, M. (2013). *Joint Commitment: How We Make the Social World*. Oup Usa.

Hedahl, M., & Huebner, B. (2018). Sharing Values. *Southern Journal of Philosophy*, *56*(2), 240–272. https://doi.org/10.1111/sjp.12275

Holton, R. (2009). *Willing, Wanting, Waiting* (1 edition). Oxford University Press.

Jaeggi, R. (2014). *Alienation* (F. Neuhouser, F. Neuhouser, & A. E. Smith, Trans.). Columbia University Press.

Khader, S. J. (2020). The Feminist Case Against Relational Autonomy. *Journal of Moral Philosophy*, *17*(5), 499–526. https://doi.org/10.1163/17455243-20203085

List, C., & Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.

McGonigal, J. (2011). *Reality Is Broken: Why Games Make Us Better and How They Can Change the World* (Reprint edition). Penguin Books.

Merry, S. E. (2016). *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. University of Chicago Press.

Millgram, E. (1997). *Practical Induction* (y First edition edition). Harvard University Press.

Nguyen, C. T. (2019). The Forms and Fluidity of Game-Play. In T. Hurka (Ed.), *Games, Sports, and Play: Philosophical Essays*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198798354.001.0001/oso-9780198798354-chapter-4

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. https://doi.org/10.1016/j.tics.2008.02.014

Owens, J., & Cribb, A. (2019). 'My Fitbit Thinks I Can Do Better!' Do Health Promoting Wearable Technologies Support Personal Autonomy? *Philosophy & Technology*, *32*(1), 23–38. https://doi.org/10.1007/s13347-017-0266-2

Porter, T. (1995). *Trust in Numbers*. Princeton University Press. https://press.princeton.edu/titles/5653.html

Reber, R., & Unkelbach, C. (2010). The Epistemic Status of Processing Fluency as Source for Judgments of Truth. *Review of Philosophy and Psychology*, *1*(4), 563–581. https://doi.org/10.1007/s13164-010-0039-7

Rose, N. S. (1999). *Governing the Soul: The Shaping of the Private Self*. Free Association Books.

Rovane, Carol. (2019). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton University Press.

Rovane, Carole. (1997). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton University Press.

Schull, N. D. (2012). *Addiction by Design*. Princeton University Press. https://press.princeton.edu/books/paperback/9780691160887/addiction-by-design

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

Strevens, M. (2013). No Understanding Without Explanation. *Studies in History and Philosophy of Science Part A*, *44*(3), 510–515. https://doi.org/10.1016/j.shpsa.2012.12.005

Superson, A. (2000). Deformed Desires and Informed Desire Tests. *Hypatia*, *20*(4), 109–126. https://doi.org/10.1111/j.1527-2001.2005.tb00539.x

Superson, A. (2010). The Deferential Wife Revisited: Agency and Moral Responsibility. *Hypatia*, *25*(2), 253–275. https://doi.org/10.1111/j.1527-2001.2010.01103.x

Velleman, D. (1989). *Practical Reflection*. Princeton University Press.

Velleman, J. D. (2009). *How We Get Along*. Cambridge University Press.

Westlund, A. C. (2003). Selflessness and Responsibility for Self: Is Deference Compatible with Autonomy? *Philosophical Review*, *112*(4), 483–523. https://doi.org/10.1215/00318108-112-4-483