Jose C. Sanchez

591 Paper

RATIONALITY AND THE CASE AGAINST PATERNALISM

INTRODUCTION

Public officials are usually in a privileged position to make the right call when compared to the rest of citizens. To begin with, it is often their professional occupation to collect the relevant information on how to achieve certain ends most efficiently, whether this be in finance, health, or any other area conceivably encompassed by public policy. Most citizens, on the contrary, cannot afford to spend time and resources to collect all this information. In addition, recent research shows that they are affected by deeply embedded biases that operate at the subconscious level and thus are unlikely to be overcome[1]. Given their more emotionally distanced position and their expertise, public officials are likely to think more rationally. While there can be doubts about the exact extent of this epistemic superiority of public officials *vis-à-vis* citizens, it seems entirely implausible to deny that it often is considerable. Therefore, it seems that in many cases citizens would be better off if public officials made decisions for them, even if this is done against their will, because they are simply more likely to obtain the results that citizens themselves want this way.

While the appeal of this paternalist position is clear and the underlying logic compelling, it is ultimately incoherent. The consequentialist case for paternalism fails on its own terms, or so I shall try to argue. My thesis will thus be of a somewhat stronger nature

[1] One of the most influential works in that field is Richard H. Thaler and Cass R. Sunstein, *Nudge : Improving Decisions about Health, Wealth, and Happiness*, Revised and expanded edition. (New York: Penguin Books, 2009).

than appealing to considerations strictly outside of the basic reasoning exposed above. I argue that this reasoning's problem lies in focusing our attention on specific cases in which it seems counterintuitive to deny that public authorities have the epistemic upper hand, as in doing so we lose track of the likely consequences of systematically implementing such policies. More careful attention to the underlying logic of the paternalist claim, and specifically to the concept of rationality, reveals that it is essentially impossible to implement a system of paternalist policies as intended because they eliminate the process that reveals the information that these policies would need to rely on to succeed in the first place. Given the forcefulness with which she makes the case for what most would consider unacceptably intrusive types of paternalism and her engagement with an (at least) equally forceful anti-paternalist, J.S. Mill, on whose argument I draw to defend my thesis, I will focus on Sarah Conly's argument in *Against Autonomy*[2].

Before beginning my argument, I specify the sort of strategy I will pursue. First, while I set aside concerns about the intrinsic harm of violating individuals' autonomies, I do not mean to imply that these are unimportant or insufficient to ground a rejection of paternalism. My argument is, in that sense, a somewhat stronger objection to the consequentialist paternalist rationale: it is *internally* inconsistent in failing to establish the relevant type of tendency to produce good outcomes. In addition, the strict consequentialist can certainly take these harms into account simply by assigning freedom of choice and autonomy some type of intrinsic value[3]. While this more nuanced evaluation of consequences would presumably still not entail a wholesale rejection of paternalism, it certainly could be invoked to reject some of the most worrisome types of intervention. Perhaps a law making it illegal to hit the snooze

[2] Sarah Conly, *Against Autonomy : Justifying Coercive Paternalism* (Cambridge: Cambridge University Press, 2013).

[3] Dworkin reads Mill's main thesis in chapter III of his *On Liberty* in a similar vein. While I later offer an alternative interpretation of the argument in that chapter, I see no reason to consider these as mutually exclusive Gerald Dworkin, "PATERNALISM," *The Monist* 56, no. 1 (1972): 64–84, https://www.jstor.org/stable/27902250.

button in the morning would make me better off if we ignore the harm to my self-esteem it causes, but the consequentialist can take this psychological harm into account and conclude that it outweighs the potential benefits. This is not the type of argument I will be pursuing. Second, I do not question that governments are often indisputably better judges than citizens, especially in matters of practical reasoning, although some would surely question this (arguably, J.S. Mill did). Third, I do not question, either, the good intentions of governments. My argument is that even if we know for a fact that the government is a better judge than the individual in a given case or a given type of cases and that it will certainly act with the intention of promoting the good of its citizens rather than in order to justify a more sinister interest on paternalist grounds, paternalist policies are not *ipso facto* justified on consequentialist grounds. Fourth, I will not pursue a "slippery slope" type of argument, either; while I do emphasize that paternalists tend to underestimate how far the consequences of paternalist restrictions can reach, I will not challenge the extent to which we can restrict the application of such legislation[4]. Finally, I assume that the background conditions are sufficiently just to make consent or lack thereof morally relevant; i.e., I argue against the paternalistic rationale *per se*, rather than against only seemingly paternalistic policies that are in reality justified (or justifiable) on different grounds (presumably on "other-regarding" grounds). My hope is that forfeiting these external concerns will focus the argument more directly on the consequentialist logic presented above itself, thereby making the argument against it more compelling. Having said that, I insist that I do not mean to imply that any of the previous lines of argument is less relevant or weighty than the one I pursue. On the contrary, one can sustain (as many indeed have) a self-standing case against paternalism based solely on any and all of the previously stated objections by themselves.

[4] Conly addresses similar objections to her own theory in chapters 4 and 5 of Conly, *Against Autonomy*.

To further clarify the structure of my argument, while I will draw on J.S. Mill's argument in *On Liberty*[5], I cannot claim to simply be restating or reinterpreting Mill's views, and it is a somewhat separate question to the one I will be pursuing in this paper whether or not Mill would have agreed with my argument or, for that matter, whether Conly's assessment of Mill's arguments against paternalism is sufficiently accurate or charitable (it might well be). That said, I do argue that my formalization of Mill's argument is plausible as such by pointing out some tensions in *On Liberty* that can potentially be resolved by my account.

I THE CONCEPT OF RATIONALITY AND PATERNALISM

Conly's defense of paternalism stems from a rejection of the classical conception of rationality. She states at the introduction of *Against Autonomy* that:

> the ground for valuing liberty is the claim that we are pre-eminently rational agents, each of us well suited to determining what goes in our own life. There is ample evidence, however, from the fields of psychology and behavioral economics, that in many situations this is simply not true.[6]

Admittedly, this rejection is indeed tempting given the obvious deficiencies of that model of rationality. For decades, neoclassical economists have used as their model the now (in)famous *homo economicus*. Such beings are perfectly rational in the sense that no cognitive limitation burdens them: particularly, they can be expected to produce calculations

[5] John Stuart Mill et al., *On Liberty, Utilitarianism, and Other Essays*, New edition, Oxford World's Classics (Oxford: Oxford University Press, 2015).
[6] Conly, *Against Autonomy*, 3.

of any level of complexity (even ones that computers struggle to simulate), and to be aware of the exact strategic environment they find themselves in, which includes all of the possible actions available to them and to every other individual that might affect their payoffs, as well as, of course, awareness of all the empirical information theoretically available to them. With all of this information at hand, *homo economicus* proceeds to predict which of his actions will produce the highest utility for her and acts accordingly (and by assumption maximizes her utility). In certain contexts (presumably, the ones economists tend to study in detail) this model might be helpful, but it certainly is not an accurate depiction of human rationality. Particularly, the more or less automatic conclusion of this model that paternalism could not possibly benefit individuals seems entirely implausible. Surely some individuals (or groups) have better information and superior cognitive capacities than others, such that if we assume their good intentions they would be able to coerce others into a utility maximizing behavior they could not have constructed themselves. Of course, the modeler *can* introduce specific limitations into the model; they can be *boundedly rational*, such that they for example can only incorporate into their predictions a limited amount of information about the past. These limitations, however, tend to be *ad hoc*, and at any rate do not seem to be able to reject the intuition that paternalism might certainly maximize the utility of less-than-optimal reasoners.

Against this model of rationality (which I will refer to as *classical rationality*), however, an entirely different conception exists. In this paper I focus on the idea of *evolutionary rationality*[7]. In the classical model, agents maximize their expected utility through their actions in a rather literal sense: they process all the relevant information *before* they make a decision and choose the action that is expected to produce the highest utility.

[7] So called because of its origins in biology, one of the earliest versions of the concept being found in Ronald Aylmer Fisher, *The Genetical Theory of Natural Selection*. (Oxford: Clarendon Press, 1930). For contemporary work that applies this concept to formal theory see, for instance, Brian Skyrms, *Evolution of the Social Contract* (New York, NY: Cambridge University Press, 1996).

This model is unhelpful because the cognitive capacities assumed are not argued for. As a consequence, any level of such cognitive capacities introduced into the model will be arbitrary and open to empirical challenge at some level. Evolutionary rationality, however, provides an entirely different model in this regard. As the most extreme assumption possible, indeed, players do not need to know anything at all about the expected payoff of different actions available to them[8]: we can, for the sake of illustration, assume that they randomize their actions in the first stage. The interesting dynamics begin after this initial decision: in the next stage, rather than going through the rational examination *homo economicus* undergoes, the amount of individuals that choose a given action depends in some way on the payoff that action produced in previous stages. We can interpret this dynamic in countless ways, and it can develop, similarly, in many different ways. At the most basic level, all that is necessary for individuals to eventually act in utility maximizing ways in this model is for them to be able to observe (in some abstract sense) the payoffs produced by different actions in previous stages, with no need for them to understand the causal mechanism that produced the payoffs (which *homo economicus* of course does need). In what follows, I attempt to offer one such intuitive interpretation of this model that, hopefully, is compatible with Mill's arguments in *On Liberty* and can ultimately successfully reject the argument that the superior epistemic capacities of governments ground a compelling consequentialist case for paternalism. In the fourth section, furthermore, I will argue that the behavioral economics alternative to the classical rationality model is unappealing and suffers from a conceptual weakness that makes it unhelpful as a tool for empirical prediction, failing to support the paternalist's claim that individuals are in some relevant sense uneducable.

[8] This is, of course, never to suggest that human beings entirely lack any predictive capacities whatsoever; in assuming the most minimal level of rationality conceivable I seek to meet the highest burden of proof possible.

As I hinted at earlier, the mechanism by which individuals update their way of behaving in search for the best outcomes does not need to be very cognitively demanding. If it were, as Conly argues, it would be hard to argue against the paternalist that governments are not often superior to individual at this task; public officials, both because of their education and their more emotionally dettached position, are indeed more likely than most individuals to identify the best means to achieve any given end. However, there are tools available to individuals that enable them to optimize their behavior far beyond what this sort of rational calculation can realistically produce. In general, it is enough for experimentation to be allowed. First, if one same individual can try different courses of action in similar circumstances over time, she is likely to identify which produces the best outcomes, even if she entirely ignores the causal process that produced them, and hence to simply repeat what worked best in the past. Actual human beings are of course not entirely ignorant of these processes; they will thus usually find the optimal courses of action even more quickly than an agent with no capacity of causal analysis whatsoever would. Second, and what is perhaps most important to Mill's argument against paternalism, the experimentation of others will serve essentially the same function, making the learning process more efficient: individuals can observe how *others* act and what their payoffs from those actions are, such that they could imitate the behaviors they perceive as optimal (i.e., those that produce outcomes they prefer the most for themselves). Note that this is not a very cognitively demanding process at all; even individuals with no cognitive capacities whatsoever can participate in it to some (admittedly reduced) extent, either by trial and error or by observing the behavior of others.

The concept of evolutionary rationality in fact finds its roots in theoretical biology. Since this story captures some of the technical details that distinguish it from the classical rationality concept in formal theory, it is worth presenting its basic structure as a way to illustrate its relevance to the present debate. Although the framework I now present is game

theoretic (i.e., based on the strategic interactions that agents are subject to when the actions of other individuals affect their own payoffs), it encapsulates the key dynamics that define the logic of evolutionary rationality as a more general decision theoretic concept.[9] I now thus turn to contrasting the classical explanation of how Nash Equilibria are derived with the evolutionary one I rely on here.

Consider a simple strategic interaction like a game of Rock, Paper, Scissors. How would I decide what to play in such a game? Presumably, I would do so by a sort of internal deduction of what sort of outcome I would expect from each of my possible choices. Upon some minimal reflection, it will be evident that I should play each strategy (rock, papers, or scissors) exactly one third of the time: if I play any one of them more often, my opponent will respond by playing the strategy that beats it more often, and I will lose more often (for example, if I always play scissors, my opponent will always play rock and I will always lose). So, I deduce, I should stick to that strategy[10].

This thought process seems completely innocuous as an explanation of how we play such a simple game. The paternalist correctly objects, however, that on a closer look it is not innocuous at all, and it does not serve to explain even slightly more complicated interactions. In a word, the explanation assumes *common knowledge*: that all players know the payoffs both they and their opponents will receive from each strategy profile and that this is known

[9] This clarification is necessary because the reader who is interested in the arguments presented here as a formalization of Mill's theory in *On Liberty*, especially as they relate to his so-called Harm Principle, might worry that a game-theoretic treatment of the concept of evolutionary rationality automatically disqualifies it as an interpretation of that principle. Specifically, the most expansive interpretation possible of what constitutes "harm to others" might hold that it would be any negative consequence an agent's actions have on another's. If this interpretation is favored, any defense of the harm principle along the lines I am here pursuing would need to be modelled as there being only one agent, thereby excluding a game-theoretic treatment altogether. It might not be obvious from my discussion, but such an interpretation could hardly be any more stringent. However, I do wish to eschew this disquisition in this paper: the purpose of the following game-theoretic explanation is only illustrative, and I stick to the one-agent decision theoretic treatment for my engagement with the paternalist argument (which, if successful, I would argue meets the relevant burden of proof).

[10] Assuming symmetric payoffs (i.e., my opponent and I stand to gain or lose the same regardless of who wins and how), this strategy is a Nash Equilibrium (the unique one in this game).

by everyone[11]. This assumption clearly begs the important questions to settle the paternalist claim that an external agent could know better than some other agent what she should do to maximize her utility. An alternative explanation of how such an equilibrium might be arrived at is thus necessary, and it should not simply assume this common knowledge condition.[12]

Evolutionary rationality becomes salient exactly in order to meet this challenge.[13] Consider the process by which species biologically adapt to their environments. The concept that typically explains these adaptations is that of an evolutionarily advantageous trait: certain traits make it more likely for specimens who possess it to produce more fertile offspring than those that do not. Over time this leads to the trait being increasingly common within that species. The trait selected for is usually a genetic mutation that happens, really, by chance. It is almost impossible that any one individual will evolve an evolutionarily advantageous trait, but in a very large population all genes are likely to experience mutations at some point, making it likely that the species will be able to adapt to new environments. Obviously the specimens that inherit the new evolutionarily advantageous trait did not plan to do so, and yet it seems that the behavior of the species as a whole resembles some form of rationality in that it is able over time to find optimal adaptations to maximize the reproductive fitness of its members.

This story, by itself, is clearly useless for the purposes of political theory (or any other social science). Biological evolution takes thousands of years and its potential effects are limited in scope. The previous was just an illustration that serves to give an intuitive overview of the type of mechanism that evolutionary rationality relies on to avoid the

[11] And that it is known by everyone that it is known by everyone, and so on.
[12] As I argued before, it should in fact ideally not assume *any* level of pre-existing knowledge of the payoffs that result from given actions: whatever such level is chosen, it will be open to empirical challenge, and thus entirely contingent for my present purposes.
[13] The concept was introduced in J. Maynard Smith and G. R. Price, "The Logic of Animal Conflict," *Nature* 246, no. 5427 (November 1973): 15–18,; while the classical reference for its formal treatment is John Maynard Smith, *Evolution and the Theory of Games* (Cambridge: Cambridge University Press, 1982).

knowledge problem I explained before. What all evolutionary processes have in common is, first, some mechanism that gives each possible strategy a non-zero probability of being tried at each iteration of the game, and, second, some mechanism by which successful strategies are more likely to be played in future iterations than unsuccessful ones. I will refer to these as *mutation* and *imitation*, respectively. There are, however, many ways such an evolutionary model can be interpreted depending on its aim. In what follows I hope that a reading of Mill's arguments in *On Liberty* through the lens of this type of model serves as one such interpretation for the anti-paternalist argument at hand. One might worry that the previous explanations left unrealistically little room for individuals making mistakes. What needs to be explained to reject the paternalist rationale I argue against here is how the presence of such mistakes does not justify paternalist interventions, which might initially seem admittedly counterintuitive. Since I do not wish to argue anything to the relatively simplistic effect that governments are even more likely than individuals to make mistakes (because, as I explained in the introduction, it is essentially impossible to argue that this would always be so), I now turn to explaining in what sense mistakes are still likely to occur even under my evolutionary analysis, and, more importantly, how this account of how mistakes are made shows that paternalist interventions could not remedy them.

We find relatively clear analogs in said work of both mutation and imitation mechanisms, and these serve to explain both why we still observe mistakes and why paternalist interventions cannot be justified on those mistakes. First, Mill's insistency on individuality, especially when understood as mere difference from the norm, makes perfect sense within an evolutionary framework. Mill's concept of "Experiments in living"[14] play the role of mutations in the previously explained framework. As such, the more experiments in

[14] John Stuart Mill, *On Liberty, Utilitarianism, and Other Essays*, ed. Mark Philp and F. Rosen, New edition, Oxford World's Classics (Oxford: Oxford University Press, 2015), 78.

living there are, the more quickly will new improvements emerge. This is so even if most of these experiments fail, which might be why he places such great emphasis on originality, even if its intrinsic valuable is questionable for the individual at times. The less individuality there is, the less likely it is that optimal behaviors will spread. Paternalist interventions, clearly, cannot possibly alleviate this source of mistakes: they limit the mutations that can happen at any given moment and thus in fact make matters worse. The second mechanism, imitation, is somewhat less straightforward to read into Mill's framework, precisely because the emphasis on originality would seem to make it undesirable to be generally imitative. However, I would argue that the process free speech plays the role of imitation in his argument. In a model of cultural evolution, unlike in a biological one, imitation is not automatic in the sense that it requires agents to be able to observe the behavior of others and to imitate successful ones. The more actions an agent can observe, the more likely she is to be able to imitate the most successful one. However, it is not necessary for agents to literally observe the behavior of others. In my reading, Mill sees in free speech a mechanism by which to maximize exposure to different ways of living, thereby making convergence to optimal ones more efficient. Here, again, is a potential source of mistakes: even if successful "mutations" do happen around an agent, she might fail to imitate them. Clearly, paternalist interventions would not remedy these mistakes from this source, either.[15]

In sum, thus, evolutionary rationality is no panacea: there can be deficiencies in both of its component mechanisms of mutation and imitation. Paternalist interventions, however, mostly limit the former, so they fail to establish the relevant tendency to produce good outcomes. The main objection the paternalist will raise against this model is that the imitation

[15] Conly's proposals would mostly fall within the first category (limits to the "mutations" that can happen) . Few theorists, to my knowledge, openly defend wide-ranging limits to free speech, so in general I would expect the second category of mistakes to be less practically relevant.

mechanism is too demanding: that individuals are, in some sense, uneducable. Conly in fact argues that individuals might either learn too late (as when their mistakes are lethal) or not be able to learn at all as a consequence of their cognitive biases. I later respond to this claim, based on the empirical research of behavioral economics, in more detail. At this point I only seek to explain how a cognitively limited agent can still act rationally in the sense of producing good outcomes.

In what follows, I discuss Mill's arguments in *On Liberty* through the lens of the evolutionary framework. The main purpose of doing so is showing how that particular reading of these arguments escapes the criticism Conly directs at it, and thus how the anti-paternalist conclusion is well supported. In addition, I hope it serves to interpret free speech as playing the role of imitation mechanisms I explained above.

II MILL'S ARGUMENT AGAINST PATERNALISM

Mill's *On Liberty* can be fruitfully understood as a broadly evolutionary account, or so I argue. I will later discuss two types of epistemic issue the paternalist project faces which I call static and dynamic, respectively. The dynamic type of problem for the paternalist relates to the production of *new* knowledge. In the model I presented, new discoveries can happen essentially by chance. Just as there is always some chance that mutations occur and as those that deliver evolutive advantages will over time become increasingly common, new discoveries can similarly occur as long as there is some chance that some individual at some point chooses a more efficient plan of action than any that was previously known. If individuals are allowed to observe this behavior, they will tend to imitate it, so that it will propagate just like an evolutionarily advantageous mutation would. The paternalist project, by definition, limits the frequency with which this can happen, as it seeks to increase the cost

or ban altogether certain actions. New ways of acting that were previously thought to be irrational in some respect will thus be less likely to emerge. It is this dynamic superiority of anti-paternalist systems that finds it clearest support in Mill's work. Both in *On Liberty* and *Utilitarianism*, Mill distinguishes his doctrine from classical utilitarianism mainly in that he considers that it has to be "utility in the largest sense, grounded on the permanent interest of man as a progressive being"[16] that provides the ultimate sanction of his ethical doctrine. This immediately turns his analysis away from direct application of the greatest happiness for the greatest number principle, searching instead for intermediary principles that will advance this end most efficiently. Conly is right, in my view, to point out that Mill seems excessively optimistic about individuals' rationality at times. However, we should keep in mind that it is mainly a concern with the fallibility of individuals that leads him to defend the mechanism of free speech. He repeatedly argues that censorship could only make sense if the censor assumed her own infallibility, which he finds a nearly metaphysical implausibility. Furthermore, Mill argues that it is not anyone's epistemic superiority that gives us reassurance of the correctness of their believes, but the fact that it is open to contestation by others:

> If even the Newtonian philosophy were not permitted to be questioned, mankind could not feel as complete assurance of its truth as they now do. The beliefs which we have most warrant for, have no safeguard to rest on, but a standing invitation to the whole world to prove them unfounded. […] This is the amount of certainty attainable lay a fallible being, and this the sole way of attaining it.[17]

[16] Mill et al., *On Liberty, Utilitarianism, and Other Essays*, 14.
[17] *Ibid*., 23.

This also explains why he finds it important to not just be theoretically aware of the arguments against one's own beliefs, but actually having to defend it from the arguments of individuals who sincerely oppose those beliefs: no one individual could rationally assess all the possible alternative views to her own. Instead, free speech enables a sort of epistemic division of labor, whereby all the possible views are explored in detailed and contrasted with others without any one individual needing to complete this immensely cognitively demanding task on her own:

> The only unfailing and permanent source of improvement is liberty, since by it there are as many possible centres of improvement as there are individuals.[18]

Mill thus seems, against what Conly tends to imply, to be very aware of the cognitive limitations of individuals. In the sense I explained, it is precisely a recognition of these limitations that leads him to place such heavy weight on the mechanism of free speech: precisely because no individual or agent could herself assume her own infallibility to produce optimal outcomes is it necessary to allow individual freedom to challenge prevailing customs to guarantee the permanent interest of man as a progressive being. On the evolutionary rationality paradigm this intuition is easy to understand: just as a species with an increased chance of mutation would evolve more quickly and efficiently, a society of humans with more individuality overall is, other things equal, more likely to find the optimal ways of acting (the more original individuals would play the role of mutations). Mill's argument, therefore, does not decisively depend on the cognitive superiority of any one agent, so that the evidence presented by Conly (and other paternalists) that specific individuals are in an inferior epistemic position regarding their governments becomes tangential. It is, instead,

[18] *Mill, 69.*

grounded on the epistemic superiority of the *process* of free speech, an argument he extends to self-regarding actions generally.

While my reliance on contemporary formal theory concepts makes my reading of Mill's thought somewhat indirect in nature, there is solid textual evidence to support such a formalization of some of Mill's most fundamental ideas in *On Liberty*. Particularly, it helps make sense of some lines of argument in chapter III that might otherwise seem to stand in tension with each other and with the other fundamental commitments of said work, if not in flat-out contradiction.

There are two (not necessarily exclusive) major lines of interpretation of *On Liberty*'s insistence on the value of individuality and its consequent rejection of paternalism. The first we might call the "instrumental" one: individuality is instrumental to the maximization of utility. This can be for a few reasons. Prominent among those, however (at least for the purposes of this essay), is the view that individuals tend to be better judges of what is best for them. For one thing, they perceive the costs and benefits of different actions directly, whereas public officials need to rely on indirect sources for this information. I have argued that this is indeed an important aspect of Mill's argument, but it is in my view unclear if this could, by itself, ground the uncompromising rejection of paternalism that Mill is after in *On Liberty* (as am I, of course). It seems open, at least in principle, to the sort of empirical challenge behavioral economists and Conly appeal to. Even if we agree with Mill that individuals are more likely to know what is best for them than public officials, we could not reject the paternalist rationale *qua* paternalist, or at least not on the purely consequentialist terms Mill relies on: there surely are exceptions to this generally superior judgment of individuals, and finding out which instances are such exceptions and which are not would presumably require similar empirical investigations to the ones behavioral economists draw on and, specifically, Conly's rejection of the point on empirical grounds could be correct. At any rate, the

paternalist will always be able to claim on contingent empirical grounds that a given policy *would* be utility maximizing, and the debate would turn on the empirical question of whether that belief is right, *not* on the soundness of the paternalist rationale *per se*. This first interpretation of the argument in *On Liberty* thus fails as a rejection of the paternalist agenda, as it collapses into a standard contingent utilitarian argument.

It in fact seems unlikely to me that this first interpretation could be sufficient for Mill's argument. The other available interpretation is, naturally, that Mill assigned some intrinsic value to individuality beyond the likely consequences of individuals choosing their actions on their own. This comes in at least two versions; a moderate one that can be read as within the previous line, and a stronger one that might place Mill in a perfectionist ethics framework altogether. The moderate alternative, so to speak, is that choosing for oneself contributes to the development of our full human faculties. Mill repeatedly insists that in choosing for herself the individual indeed employs all her faculties, as opposed to merely engaging in the "ape-like" faculty of imitation[19][20]. We could read this as suggesting that even if an individual is likely to make a mistake in choosing an action for herself, the long-term consequences of allowing her to make that mistake are likely to be beneficial in improving her faculties and with them her tendency to make the right choice. As might be obvious, however, this argument would be open to the same challenge I outlined in the previous paragraph: whether this is so or not is, in short, a contingent empirical matter. Admittedly, it is more difficult to empirically verify such a long-term prediction, and this difficulty might be

[19] This famous quote might be read to contradict my evolutionary interpretation, which relies on the mechanism that I myself termed of "imitation". Such a reading would be wrong, in my view: Mill means by "imitating" the mere reproduction of others' actions irrespective of the appropriateness of doing so to one's nature and circumstances, or more broadly regardless of the expected contribution to one's own utility. I use the word "imitation" in a much more specific sense, as is hopefully clear from the rest of the argument in this essay: perhaps most importantly, individuals update their behavior based precisely on their expectation that imitating other individuals' actions would make *them* better off than continuing with their current behavior. Mill uses the word "imitate" in its conservative (in terms of behavior) meaning, whereas I very much use it in a progressive connotation.

[20] Mill, *On Liberty, Utilitarianism, and Other Essays*, 58.

invoked to reject paternalist policies in as many instances as the evidence (or lack thereof in) supports, but the argument fails as a theoretical rejection of the paternalist rationale *per se* on the same grounds as the previous one does. Both, therefore, seem to me to be implausible readings of Mill's argument in *On Liberty*, which should give us pause before accepting them as the fundamental pieces of that work (as Conly likely does).

Perhaps more initially plausible is the stronger version of this argument, which is that individuality has intrinsic value *per se* regardless of whether its exercise and development in fact contributes to "the greatest happiness for the greatest number" as usually conceived or not. Mill's language in the following quote makes it tempting, to say the least, to settle on this interpretation:

> He who chooses his plan for himself, employs all his faculties […] It is possible that he might be guided in some good path, and kept out of harm's way, without any of these things. But what will be his comparative worth as a human being? It really is of importance, not only what men do, but also what manner of me they are that do it[21]

While it is certainly plausible to interpret this line of reasoning of Mill to contradict my reconstruction here, I do not in fact think that this tension turns out to be fatal to it: as long as Mill's argument can be read as a broadly consequentialist one, my interpretation should remain equally meaningful. The perfectionist or virtue ethics component Mill seems to be introducing in this part of *On Liberty* can, logically speaking, certainly fit into a consequentialist framework, even if not necessarily a utilitarian one. One would simply have to recognize that individuals intrinsically care about individuality, and that paternalist interventions naturally pose an obstacle to its development.

[21] Mill, 58.

Regardless of these considerations, however, my main aim in this section is, in fact, to move beyond these alternative interpretations and suggest that the evolutionary interpretation of Mill's anti-paternalist argument I have provided and defended here can resolve the tension between chapter III and earlier parts of *On Liberty*. If so, it potentially offers the most parsimonious interpretation of that work, which would (other things equal) speak in favor of it not just as a piece of independent normative argument, but also as an actual reconstruction of Mill's argument.

The main problem with the lines of interpretation I just briefly reviewed is that Mill himself insists that he regards "utility as the ultimate appeal on all ethical questions"[22] and declares both in *On Liberty* and elsewhere (including, obviously, his *Utilitarianism*) that he remains loyal to "the greatest happiness for the greatest number" principle. In light of this commitment it becomes unclear how to make sense of chapter III of *On Liberty*: surely both one's use or misuse of her faculties and her "comparative worth as a human being" are independent of the advancement of happiness, at least in principle.

My evolutionary reading of his argument, however, shows an appealing way of solving this tension. Following Mill's defense of the intrinsic value of individuality in chapter III of *On Liberty* we in fact see what I think are the passages that support the evolutionary rationality model I defend here the most clearly.[23] The similarity between what Mill calls "originality" or "genius" and the concept of a mutation in an evolutionary process is uncanny: both are rare deviations from the usual strategies ("customs" as Mill calls them) that might initially seem pointless, but that enable progress and prevent human life from becoming "a stagnant pool"[24]. Crucially, furthermore, these innovations are not only of use to the innovators themselves, but to the rest of society. This brings home, in my view, the

[22] Mill, 14.
[23] See for instance the reference in footnote 18 above
[24] Mill, *On Liberty, Utilitarianism, and Other Essays*, 63.

evolutionary understanding of the argument of *On Liberty*, and with it makes the intrinsic value argument in chapter III compatible with the overall aim of the work: what Mill defends in *On Liberty* is that individuals' liberty to experiment and to simply make mistakes is a necessary precondition for a collective system that enables progress and the fine-tuning of sub-optimal ways of acting, and thus should be allowed. Experiments are valuable because of this, even if they go wrong, and even if they seem reckless from the perspective of an outsider. Insofar as their negative consequences for the individual are thus not a reason to prevent them, they can be seen as intrinsically valuable in this specific sense. The preference for individual liberty, however, is ultimately grounded on utilitarian grounds, together with the evolutionary understanding of the collective nature of progress. The tension between the two interpretations arises when one assesses the argument from the perspective of classical rationality, but vanishes when one considers the evolutionary alternative, and with it the inherently social nature of knowledge and progress. Like this, the parts of *On Liberty* that seem to contradict the consequentialist rationale of the rest of the work fit into a theory that is utilitarian overall when read as a defense of the value of experimentation in an evolutionary process, in my view lending such an interpretation considerable plausibility.

As a final remark, Mill comes back to a different version of the instrumental argument of individuality at the end of chapter III: now the focus is not on the superior judgment of individuals about their own well-being, but on a the degree of variance of individual natures and preferences that make different actions optimal for different individuals. Part of his worry here is that societies tend to become more homogenous over time in this respect. It is easy to see why this would be a problem from the evolutionary perspective: if indeed individuals become more alike, mutations or experiments in living become more unlikely, and with it progress slower and more inefficient. Defending individual liberty within the limits of the Harm Principle would thus become increasingly important. The reason I highlight this

interpretation is that this argument might seem inconsistent with the rest of the work were it not because of the evolutionary interpretation I propose: as individuals become more alike, we might think that the range of strategies (in formal language) they should follow should correspondingly narrow down, and fail to see this as a problem (and hence be more likely to support paternalist policies). Within the evolutionary framework, this lack of experimentation would be equivalent to a stagnation of human progress.

In the next section, I examine two types of epistemic challenge that, I argue, are a fatal problem for the paternalist rationale, as an evolutionary framework shows.

III THE INFORMATION PROBLEM OF PATERNALISM

The rest of this essay explains how, given the nature of the epistemic superiority of a system in which individuals are free in self-regarding actions, the paternalist argument fails to establish that banning certain presumably harmful choices by individuals could tend to produce better outcomes.

My previous discussion might suggest a chronological conception of the epistemic inferiority of paternalist systems: i.e., that they are inferior because they prevent innovation over time from occurring. This is undoubtably part of the argument. I return to this facet below, and it might well be the most significant objection to paternalism. However, I argue that the paternalist argument fails to establish the relevant epistemic superiority even if analyzed as a "static" case; i.e., even if we ignore the *future* epistemic losses that are likely to be caused by such a set of policies.

Conly (and paternalists, in general) underestimates the amount of knowledge that is required for public officials to make good decisions for specific individuals. As I argued

before, she seems to have a model in mind in which individuals have some goal or end that they are aware of desiring, but simply fail to identify the right means of obtaining it. Gerd Gigerenzer[25] describes this model that paternalists tend to imply by analogy to visual illusions: there exists an objectively discernible object that can be better observed by the paternalizer (the public official deciding the paternalist policy) than by the paternalized subject, who might display a cognitive bias akin to an optical illusion that prevents her from correctly identifying the object. One can only hold on to such a view if, as Conly does, she assumes a clear distinction between means and ends (in the metaphor, objects). She expresses this point explicitly in chapter 3 of *Against* Autonomy:

> There is a clear difference between a paternalism about ends and a paternalism about means, and that difference means that much of the fear that individual lives will be subsumed under some foreign system of values is, in this case, not applicable.[26]

This distinction, however, is largely artificial. Different ways of producing seemingly equivalent outcomes involve tradeoffs *of different ends*. Conly for example argues that cigarettes should be banned. Scientifically, we can establish that it is unhealthy to smoke. We can also safely assume that most people value health. Conly would thus have it that deciding to smoke constitutes a failure of instrumental reasoning, but this surely does not follow: smoking *does* provide benefits, obviously including immediate pleasure, and it is entirely conceivable that this could outweigh the costs for the individual. For the paternalist policy to not involve the moralist judgments about which ends individuals ought to desire and to what extent that Conly seeks to avoid, the paternalizer would need to know, at least roughly, the

[25] Gerd Gigerenzer, "On the Supposed Evidence for Libertarian Paternalism," *Review of Philosophy and Psychology* 6, no. 3 (September 2015): 361–83, https://doi.org/10.1007/s13164-015-0248-1.
[26] Conly, *Against Autonomy*, 89.

relative cardinal value attached to different ends by the individual. It would not be enough to be aware of just how much the individual values the main end pursued by the paternalist policy, but it would also be necessary to know whether the involved tradeoffs of other ends would be compensated by the (presumed) gain. It is thus not enough for the paternalist to be a financial expert in order to successfully decide how much debt an individual should acquire; they also need to know just how much the individual discounts future consumption as compared to present (which in turn depends on a myriad of factors), and even how much she enjoys the process of making financial decisions herself, just to mention a few parameters. Furthermore, no decision's consequences begin and end with the decision at hand: they instead modify the tradeoffs that will be involved in other choices. Smoking is a particularly clear example of how the paternalist policy could in fact hardly accomplish its goal: in Conly's view, there seems to be a failure of self-control causing the instrumental reasoning mistake of smoking. Self-control, obviously, cannot be mandated: the individual that craves cigarettes but is not allowed to consume them will most likely compensate through some means not contemplated by the policy; she might consume other drugs, or fatty foods, etc.

By itself, the previous need not be a theoretical objection to the paternalist policy, but a merely pragmatic one: we might think that, challenging as it is, the potential paternalizer might still be more likely than the individual to understand all the necessary information, and thus to make a better informed policy. The claim becomes far more unlikely, however, if we consider the previously explained rationale for the epistemic superiority of a non-paternalist system.

Namely, the individual does not need to be consciously aware of all the tradeoffs involved; it is enough for her to be allowed to try alternative actions for her to simply perceive which leaves her better off all things considered. This alternative is not available as a matter of paternalist policy. In order to arrive at the optimal decision, the paternalizer does

need, in theory, to consciously and rationally be aware of the tradeoffs involved, even before they are experimented by anyone at all[27], because she does not perceive the end result of the policy like the paternalized agent does; i.e., she does not *perceive* unequivocally whether the end result is optimal all things considered (more precisely, she does not know the exact magnitude of the benefit or harm caused by the policy). Even if she is far superior to the individual at a cognitive level, she lacks access to the mere trial and error the individual can engage in. It is the superiority of *this* tool, not anyone's epistemic capacities *per se*, that grounds the consequentialist case against paternalism.

At this point, furthermore, Conly has not yet explained what the alleged benefit of governmental paternalism is. If indeed it is the case that it is mere practical reasoning (i.e., about ends to means) that individuals are deficient at, they would presumably happily voluntarily delegate such decisions. They might hire personal trainers, nutritionists, financial advisors, etc. These different agents can efficiently assist them in adopting the rational (in the sense Conly uses the word) policy. In addition, these would presumably be subject to some sort of competition with each other, which would give them an incentive to maintain a good reputation that would presumably be sustained by indeed promoting the good of their clients. This possibility does not even seem to be considered by Conly, as she implies that being forced not just to pursue a specific policy, but to altogether lack the autonomy to decide whether one delegates the decision or not, has some benefit, and even seems to believe that having the option to actively opt out would constitute some sort of harm.

A natural concern about my previous argument is that, at the end of the day, one might think, we do know that some options are just bad: one could only make certain choices

[27] Some similar mechanism might be available to the paternalizer if opting out of paternalist policies is allowed. This would of course barely count as a paternalist policy. Conly in fact clearly specifies that the weaker form of paternalism Sunstein and Thaler call "libertarian paternalism" that (similarly to the caveat I just introduced) does allow opting out is insufficient in her view; she namely endorses what she calls "coercive paternalism" as a stronger type of policy that does not allow opting out.

by mistake, in a sense. I have argued before that being certain of this is more complicated than might first seem because of the interdependency of choices and the efficiency with which mere trial and error processes contribute to finding optimal ways of acting. Even if these problems were overcome, however, there is still a problem that paternalism would face, and that Conly, in my view, neglects. Our body of knowledge is in constant evolution: no one, by definition, can predict what the most efficient way of doing things will be in the future. Again, this innovation occurs as a result of a process of trial and error in which all of the individuals composing a society can engage in: no government could, by itself, reproduce this. This dynamic information problem is even more theoretically challenging for the paternalist than the previous. It is conceivable, for example, that neuro-science might produce technology that enabled the paternalist to access individuals' minds so as to observe, much as they themselves would, whether a paternalist policy left them better off or not (to empirically observe their level of happiness, or something along those lines). It is not possible, however, for them to take into account alternative policies that rely on knowledge that has not been invented yet. The previously explained process of trial and error, as Mill argues, enables individuals to advance knowledge in a more efficient way than any one agent could. Many innovations, especially those connected to morality, were once thought to be objectively irrational ways of acting, and the paternalist could have found it beneficial to ban these ways of acting. Conly would presumably not want to ban those innovation that produced good outcomes, but of course this benefit of hindsight is not available for future innovations. While one might of course find good other-regarding reasons to justify these restrictions, I have argued that allowing experiments that seem irrational produces an epistemically more efficient system than any alternative that relies on the epistemic capacities of any one agent (individual or collective).

In sum, in this section I argued that the paternalist project faces a more challenging information problem than Conly acknowledges, and that the argument against paternalist policies based on the evolutionary conception of rationality fares better against these challenges, both in their static and in their dynamic dimensions.

IV BEHAVIORAL ECONOMICS

My argument so far relies on learning (imitation) mechanisms: agents observe the results of alternative ways of acting and imitate those that produce the best outcomes for themselves. Conly (and other paternalists) would likely respond that these mechanisms are too demanding for individuals; that individuals are, in that sense, uneducable[28]. Behavioral economics is the scientific basis of that belief: its central tenet is that individuals are subject to biases that make them act against their true interests. The seminal work in this field is, of course, that of Sunstein and Thaler[29]. Many issues immediately arise with this view. However, recall that I forewent most of them; I assume that governments indeed have the altruistic intention of promoting the good of its citizens and that indeed its epistemic capacities are superior to those of any one individual (or group agent) separately. Still, I have implied in the previous that behavioral economics fails to justify a paternalist approach to policy making. In this section I explicitly state the reasons why this failure occurs. Two types of weaknesses undermine this theory: first and foremost, a conceptual one in its flawed conception of rationality and, second and as a consequence, the empirical evidence alluded to fails to justify the practical recommendations endorsed by paternalists.

[28] Conly for instance argues that individuals often learn too late to amend their mistakes, or that their failures are sometimes lethal. She also finds the view that individuals generally are likely to improve their situation if left to their own devices implausible, as I discussed on the previous section on Mill's arguments against paternalism.
[29] Thaler and Sunstein, *Nudge*.

First, one can only establish empirically that an individual failed to act in her own best interest if she knows beforehand what that best interest is. The behavioral economist and her paternalist followers reject the view that it is choices that reveal preferences, but do not tell us what alternative source of information they draw on to produce this required knowledge[30]. They simply assume that it is evident, for instance, just how much debt it is reasonable to assume, how much long term health is worth to the individual relative to immediate pleasure, etc. As I have argued, however, the challenge is not to show that our intuitions are right in any one specific instance, but that a public official or public agent could *consistently* come up with this information. In a non-paternalist system, I have argued, no one actually needs this information in its entirety: the process of trial and error itself enables individuals to maximize their utility without necessarily engaging in the sort of calculation that would require that information.

Second, given the open-ended nature of preferences (in that we do not usually know exactly why individuals value what they value), it is nearly impossible to provide empirical evidence to the effect that individuals act irrationally: there simply is no specific counterfactual. The empirical evidence appealed to in behavioral economics is no exception. For example, Sunstein and Thaler famously show that merely changing the default option in savings plans has a considerable effect in the option individuals choose, even when they are allowed to opt out of the default[31]. The conclusion of this finding, they imply, is that these individuals could not be acting rationally in the traditional sense: they reveal different preferences on the same choice set. A favored explanation of this sort of failure for the

[30] Similar points are made in Mario J. Rizzo, *Escaping Paternalism : Rationality, Behavioral Economics, and Public Policy*, Cambridge Studies in Economics, Choice, and Society (Cambridge, United Kingdom ; New York, NY: Cambridge University Press, 2020). He does not, however, provide a fundamentally different conception of rationality as I attempted to, and instead relies on a similar conception to what I called classical rationality.

[31] See for instance their article Cass R. Sunstein and Richard H. Thaler, "Libertarian Paternalism Is Not an Oxymoron," *The University of Chicago Law Review* 70, no. 4 (2003): 1159–1202.

behavioral economist is the concept of a cognitive bias. Just as someone with myopia fails to properly see an object at a distance, the thought goes, individuals can fail to correctly process information and decide a corresponding course of action when trivial factors in its presentation are changed (e.g., when the default option is changed). In my view, this model is too narrow to accurately describe the rationality that humans are capable of. An important concept behavioral economics research fails to distinguish from a cognitive bias is that of a heuristic: when an information environment is complex, individuals will rely on simpler pieces of information that they *can* process to infer the rest of the information that might be relevant to their decision. One reason this distinction is both important and challenging to the behavioral economics claim that individuals are (in some respects) uneducable is that it seems extremely complicated to produce externally valid experiments to test their interpretation. This is because empirical research on rationality almost by definition presents artificial information environments that could not reproduce the circumstances surrounding complicated (and important) decisions faced by individuals in the real world. For the paternalist to show that their plans can produce better outcomes than leaving individuals to their own devices would, they need to provide evidence of such a tendency in the actual environment the policies would be applied to.

I argued before that one can only somewhat artificially compartmentalize decisions: choices made in one regard affect the tradeoffs faced in others. In order to show the optimality of a paternalist restriction, we would thus need to analyze the overall effect for the individual (i.e., considering also the consequences beyond the decision at hand) of using the heuristic being tested. The paternalist would face a challenge in even producing a testable hypothesis: merely establishing the consequences we would need to examine is itself a cognitively unfeasible task (let alone collecting that information). To give an example; in providing empirical evidence about the negative health effects of trans fats, the paternalist has

implicitly already decided that potential benefits in other areas, such as short term pleasure, are to be disregarded. Since the paternalist project, at least in Conly's variant, does not seek to impose goals on individuals, the paternalist thus needs to know whether this benefit is outweighed by the cost or not. But she also needs to know if, say, the subject will look for other sources of pleasure if prevented from consuming trans fats, and if this will affect their work discipline, which in turn might affect their overall income, which in turn might affect their career choices, etc. I argued that as long as the individual is allowed to experiment with different choices, she can incorporate this information into her decision making without being aware of the actual causal processes that produce these outcomes. Insofar as the paternalizer lacks access to empirical perception of the end result, the information she needs to rationally assess is far greater.

In sum, if we indeed stick to the conceptual framework in which the claim that individuals choose according to their preferences is a tautology, there is no amount of empirical evidence that could refute the claim, because it is a conceptual tool, not an empirical conclusion. What a convincing rejection of this framework needs to offer is an alternative model and, with it, presumably, an alternative source of information about individuals' preferences. I have hitherto argued that such an alternative source seems intuitively unnecessary in isolated cases, but that this compartmentalization is unavailable: the burden that an alternative model to assess rationality would need to meet is thus significantly higher. There will always be different interpretations available of the very same empirical evidence depending on the model of rationality one applies. Thus, there is no way of falsifying the claim that a paternalist restriction indeed left an individual better off than she would have been without the restriction in place. Behavioral economics research fails to differentiate irrational behavior in the form of a cognitive bias from perfectly rational one in the form of heuristics because they present empirical evidence where a conceptual argument

is called for. Paternalists reject the view that individuals' actions reveal their preferences, but simply do not tell us what reveals them instead. Unsurprisingly, these conceptual issues have led to considerable challenges to the empirical evidence alluded to by Sunstein and Thaler, providing many examples of the sort of confusion I argued behavioral economics entails. For example, Gigerenzer[32] shows that individuals often interpret the default option in savings plans as a recommendation: if so, it could certainly be an efficient heuristic to use by cognitively limited agents. Their empirical evidence can hardly debunk this sort of alternative interpretation.

Sunstein and Thaler claim at the beginning of their article that the view that individuals act in order to promote their own good is either a tautology and therefore uninteresting or a testable claim, and that they favor the latter[33]. What is entirely wrong is the implication that considering it a tautology (as I defended we should do) is somehow misguided. As I tried to show in this paper, one can certainly adhere to the tautological view of rationality in a consistent, conceptually illuminating way.

CONCLUSION

I started by acknowledging that governments are often (maybe even usually) epistemically advantaged regarding individuals. If so, the paternalist view that we ought to compel individuals to do certain things or remove certain options by force from them simply seems intuitively right: why insist on autonomy when this will leave individuals worse off? In response, I argued that the information problem for governments is of an entirely different nature than it might superficially appear to be. More careful consideration of the concept of

[32] Gigerenzer, "On the Supposed Evidence for Libertarian Paternalism."
[33] Sunstein and Thaler, "Libertarian Paternalism Is Not an Oxymoron," 1163.

rationality, furthermore, lends support to Mill's arguments against paternalism: a non-paternalist system enables individuals to maximize their utility without themselves having to consciously assess all the required information to make optimal decisions simply by observing the behavior of others and the results it produces (or by experimenting with new behaviors themselves). Governments lack access to direct experience of this net result, so the same mechanism is not available to them. While the field of behavioral economics correctly highlights fundamental flaws in the classical economics' concept of rationality, it fails to provide a coherent alternative, and specifically fails to provide empirical evidence that individuals cannot learn in the way the evolutionary model requires for them to maximize their utility over time. On the contrary, paternalist restrictions limit the availability and transmissibility of relevant information, causing a considerable epistemic loss for society. As a result, I argued, the consequentialist case for paternalism is unconvincing. I reiterate, however, that my argument here is not meant to dismiss normative issues raised by paternalist policies, which themselves could well suffice to refute such proposals. The arguments in this essay are meant to apply to strictly self-regarding actions; whether these exist at all or not or to what extent is a separate question that should not affect my criticism of paternalism. It is only under the assumption of their existence that the paternalist case would be relevant at all, and it is strictly on that kind of action that I focused.

References:

Conly, Sarah. *Against Autonomy : Justifying Coercive Paternalism*. Cambridge: Cambridge University Press, 2013.

Dworkin, Gerald. "PATERNALISM." *The Monist* 56, no. 1 (1972): 64–84.

Fisher, Ronald Aylmer. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930.

Gigerenzer, Gerd. "On the Supposed Evidence for Libertarian Paternalism." *Review of Philosophy and Psychology* 6, no. 3 (September 2015): 361–83.

Mill, John Stuart. *On Liberty, Utilitarianism, and Other Essays*. Edited by Mark Philp and F. Rosen. New edition. Oxford World's Classics. Oxford: Oxford University Press, 2015.

Rizzo, Mario J. *Escaping Paternalism : Rationality, Behavioral Economics, and Public Policy*. Cambridge Studies in Economics, Choice, and Society. Cambridge, United Kingdom ; New York, NY: Cambridge University Press, 2020.

Skyrms, Brian. *Evolution of the Social Contract*. New York, NY: Cambridge University Press, 1996.

Smith, J. Maynard, and G. R. Price. "The Logic of Animal Conflict." *Nature* 246, no. 5427 (November 1973): 15–18.

Smith, John Maynard. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press, 1982.

Sunstein, Cass R., and Richard H. Thaler. "Libertarian Paternalism Is Not an Oxymoron." *The University of Chicago Law Review* 70, no. 4 (2003): 1159–1202.

Thaler, Richard H., and Cass R. Sunstein. *Nudge : Improving Decisions about Health, Wealth, and Happiness*. Revised and Expanded edition. New York: Penguin Books, 2009.