Boosting Agency: Autonomy-Promoting Behavioral Public Policy

Till Grüne-Yanoff

Royal Institute of Technology, Sweden

Email: gryne@kth.se

Abstract

Behavioral public policies (BPPs) seek to change people's behavior in some beneficial way. In this essay, I examine the effect of BPPs on personal autonomy. Categorizing BPPs by the mechanisms through which they operate, I argue that their effect on autonomy depends on their mechanism type. I investigate both the autonomy-preserving and the autonomy-promoting properties of these types. I conclude that bias-harnessing BPPs might violate autonomy while competence-enhancing ("boosting") BPPs never do; and that bias-harnessing BPPs cannot promote autonomy, while boosting BPPs can.

Keywords: behavioral public policy; boost; nudge; mechanism; agency; autonomy

Introduction

How to get people to consume energy more sustainably? One of the answers we tested in an experiment last year in Stockholm was to provide inhabitants with microwave recipes. Microwaves are great for cooking eggs, rice, many kinds of vegetables, and much more.

These dishes are tasty, healthy and quickly made, and they also save a fair amount of energy. Microwaving rice, for example, takes up to 30 percent less energy than cooking it on the stove.[1] Baking a potato in the oven consumes about nineteen times more energy than microwaving it.[2] Many people, however, are not aware of these differences, nor do they know how to cook a proper meal in a microwave. Teaching recipes provides people with competences that allow them to exploit the considerable energy-saving potential of microwaves.

Our experiment focused on the comparative effectiveness of different behavioral public policies (BPPs), which are interventions that seek to influence behavior based on scientific evidence, without coercion or incentivization. We pitted boosts (interventions aiming to increase competences, like our distribution of recipes) against BPPs that seek to influence behavior through feedback, goal commitments, or social comparisons.[3] We could show that boosts were more effective than those other BPPs for this task.[4] In the present essay, I expand this comparison to questions of personal autonomy, arguing that boosts are preferable to other BPP types because they foster agency and, through that, promote people's autonomy.[5]

The effects of BPP on autonomy have been widely discussed; the diagnosis that a nudge infringes on autonomy is often considered an ethical reason against employing it.[6] In the ongoing debate there is disagreement over how severe such intrusions are, how often BPPs compromise autonomy, and whether BPP benefits might outweigh potential autonomy infringements.[7] This debate treats autonomy largely as an absence: of violation, of illegitimate intervention, of disturbance of what is inherently there.

I want to give this debate a more constructive twist. My focus is on what to do in order to foster autonomy rather than on what not to do in order to protect it. The basic intuition is that people become more autonomous when their choices are produced by

competent deliberation that is sensitive to their reasons. Such competences are often local, simple, and rather mundane, such as the ability to prepare tasty meals in a microwave. Without such know-how, agents often cannot connect their choices to their reasons for acting in a rational way. Providing this know-how allows them to competently deliberate to make a choice, and also allows them to reflect better on the reasons they have for acting.

I have two goals in this essay. First, I want to argue for boost interventions on autonomy grounds. Boosts do not reduce autonomy, while other BPPs might. Furthermore, they foster agency and, through that, promote autonomy. Those are important pro tanto reasons to prefer boosts over other BPPs. Second, by comparing different BPP types, I will investigate underappreciated aspects of the autonomy notion itself. Autonomy is too often seen as a negative threshold concept: unless some necessary conditions are satisfied, people are said to lack autonomy. Here, instead, I seek to elaborate on ways in which autonomy can be promoted rather than merely protected.

The essay is structured as follows. The next section reviews the extant criticisms of BPPs as autonomy-reducing. This sets a minimum challenge for boost approaches: they at least need to avoid these criticisms to be preferable to other BPPs on autonomy grounds. The following section sketches the mechanistic basis for a BPP typology. On this mechanistic basis, I show that nudges need not satisfy the minimal autonomy criteria, while boosts necessarily do. The section after that goes beyond the minimal autonomy criteria and argues that autonomy is promoted by improving agency. Agency consists in the right productive relationship between reasons, reasoning, and action. Agency promotes autonomy if the thus-fostered goals are non-alienating. Finally, I show how boosts, in contrast to other BPPs, can be autonomy-promoting in this sense.

Some BPPs reduce autonomy

One of the chief criticisms leveled against many BPPs is that they compromise their targets' autonomy. These criticisms have been extensively documented.[8] I distinguish three kinds of criticisms, which serve as minimum conditions that BPPs must satisfy to be autonomy-preserving.

One worry is that BPPs might reduce the authenticity of desires or values that motivate actions: "[They] pull our strings and employ tricks to get us to do what *they* want."[9] This is not an issue of freedom of choice; most BPPs are characterized as not removing or altering options.[10] Rather, this worry concerns undue influence on how the choice is generated,[11] and ways of manipulation that "pervert choice."[12] For example, if I stick with a green utility company default, introduced by a policymaker hoping to make my energy consumption more sustainable, I might spend my budget on greener but more expensive energy, while no money remains to give to causes that I actually champion. Consequently, the thus-influenced actions might not reflect the genuine ends of the agent; the BPP then contributes to the alienation of the agent from her own desires. Such alienation is widely seen as a hallmark of heteronomy, even if views differ on how one is supposed to identify what the genuine or authentic desires of a person are.[13]

Another worry is that BPPs interfere with people's capacities for rational thought, deliberation, or reflection. A BPP might consist in "another person's willful exploitation of those less than fully autonomous … patterns of decision-making."[14] Critics who worry about the manipulation of choice, as in the quotations above, express a related idea: even if a BPP gets me to do something that serves my authentic desires or values, the fact that the BPP manipulated me (instead of rationally persuaded me) raises autonomy worries. For example, when the Save-More-Tomorrow (SMT)[15] intervention makes people choose retirement-saving deductions for next year (instead of for now), it might induce them to save more. That's because it exploits the well-known fact that many people discount the future

hyperbolically: they prefer a larger, later reward to a smaller, earlier one, until they are sufficiently close to the smaller, earlier one, when their preferences reverse. SMT changes people's choice architecture—as it's typically an externally triggered choice, not a change people themselves elect—gets them further away from the smaller, earlier reward; and hence increases the chance that they choose their preference for the larger, later one. This might be an effective intervention and the resulting choice might further people's genuine desires, but people thus influenced do not reach the practical conclusion through coherent reasoning. The BPP does not lead people to acknowledge the time-dependence of their preferences, nor does it encourage them to reflect on the comparative merits of these preferences. Such absence of self-reflection and procedural independence is widely seen as a hallmark of heteronomy, even if it remains controversial how precisely to specify what constitutes such independence.[16]

Another worry, related but distinct from the previous one, is that BPPs often are not transparent to those affected.[17] For example, when reframing a nightlight ad as "saves 20 percent electricity in comparison to its competitor" instead of "consumes 1W and its competitor consumes 1.25W" induces me to buy it, I might well have reflected and deliberated about this purchase, but I might not be aware that this framing switch triggered that purchase, which otherwise I might not have made. Consequently, the BPP obscures the reasons for my action. While I think that I bought it for its substantial energy savings, it really is the reframing as a relative increase that does the work. If I had known that the monthly energy saving from running the nightlight 24/7 was less than the energy needed for boiling a liter of water, I wouldn't have bought it. The inability to understand the reasons for one's actions is widely seen as a hallmark of heteronomy, even though what the relevant notion of "having reason to" is, and its relation to desires and interests, remains in debate.[18]

I consider these worries to be necessary conditions to address, as each of them individually indicates that a BPP violates autonomy. Some defenders of such interventions

have sought to respond to such worries individually. For example, some authors have argued that BPPs sometimes promote autonomy, in particular, when they help an agent "choose according to her own conception of the good in situations where she would have previously done otherwise."[19] While such arguments might assuage alienation worries, they often do not meet transparency and procedural independence worries. "Helping to choose" through a BPP often means to harness biases that are neither transparent to the affected subjects nor encourage them to self-reflect and deliberate. For the same reason, claiming that autonomy is not undermined as long as people's ends are respected fails to convince. Autonomy is not only about people achieving (non-alienating) ends. The process also matters.

Others have argued that these three conditions turn autonomy into a utopian goal that few humans ever reach: the non-alienation condition requires too high a level of self-knowledge and procedural independence requires too much of human rationality: "[W]e should treat people as they are. If people are less rational than they think or wish they were be [sic], then respecting them as agents does not imply treating them as perfectly rational choosers."[20] I disagree with this last claim, as any meaningful normative claims diverge from how people actually are. Instead, meaningful norms make demands that often differ from actuality, but stay within what can be reasonably expected from people, and do not go beyond what humans are capable of. I believe that the three autonomy demands, if properly formulated, meet such an ought-implies-can constraint. To that end, let me clarify some core features of the autonomy concept I have in mind.

First, I take autonomy to be local and domain specific. One might be autonomous with respect to one's diet choices but fail to be autonomous with respect to one's financial choices at the same time. Furthermore, autonomy can come in degrees. There might not be any maximum or minimum level of autonomy, but people will be more or less autonomous in certain domains, depending to what degree they meet the conditions of non-alienation,

procedural independence, and self-transparency. So even if full global autonomy might be super-human, or not even properly defined, these conditions still allow specifying whether a BPP substantially decreases the degree of autonomy that a person enjoyed in a particular domain.

Also, non-alienation and procedural independence are not as demanding as they might appear. One might be worried that they exclude any form of habituation or automatic processes, implying, for example, that my thoughtlessly setting the shower tap to a particular setting every time I shower fails those conditions. But that need not be the case. As I will argue below, acting intentionally and for reasons is compatible with automatic or habituated processes under fairly permissible conditions. Certainly, one need not consciously have reasoned oneself to an action for it to be autonomous. For example, temptation bundling trains the habit of consuming "temptations" (for example, television series) only in connection with activities one finds difficult to motivate oneself to do otherwise.[21] When the training succeeds and people habitually consume their temptation at the gym, their choices are still autonomous. As long as the subject intentionally habituated herself in order to achieve this result, the subsequent automatization does not reduce autonomy; when habituating, she had reasons for it and reasoned her way toward it. I therefore think that these conditions are not unreasonable as a minimal requirement for BPPs that aim to be not autonomy-reducing.

Boosts meet these minimal autonomy criteria

In the previous section, I reviewed several autonomy criticisms against BPPs. These criticisms do not apply to all BPPs. Instead, one would have to decide on a case-by-case basis whether a particular BPP compromises autonomy or not. This is further complicated by the fact that many of these criticisms concern the process through which the BPP intervention

affects their target. Consequently, instead of deciding whether an intervention—for example, a default-setting, framing, or temporal repositioning—is autonomy-violating or not, one needs to decide whether the intervention in a given context operates through a process that reduces autonomy.

To simplify this assessment, I am offering a typology of BPPs based on the mechanisms through which they operate. Such a typology, I have argued elsewhere, is useful to ex ante predict the effectiveness of BPP types in different kinds of contexts.[22] Here, I argue that such a typology also helps delineate the autonomy-reducing or -enhancing properties of different BPP types.

I will start with an abstract and highly simplified mechanism scheme of decision-making (see Figure 1). Individual decision-makers distinguish a number of alternatives, identify their relevant properties (for example, their possible consequences and the uncertainty with which they come about), and choose one of the alternatives according to some selection rule. This rule might involve the evaluation of consequences and their uncertainty, but it might also be a rule as simple as "choose the alternative highest up on the list."

But how do individuals arrive at a list of alternatives and their relevant properties? They search the environment according to some search rules that tell them what information to focus on, how to mark distinctions, and when to stop searching. This also means that individuals consider only some features of the environment relevant, while ignoring others. The search rule therefore distinguishes relevant information from irrelevant context. What behavioral research has shown, however, is that context considered irrelevant by individuals might nevertheless have causal impact on their decision. Some such contextual factors, for example, "anchors" or "frames," purportedly influence the representation of alternatives and their properties. Other factors, such as "default effects" or "reference points," purportedly

influence the evaluation and selection of alternatives, although the individual considers them irrelevant.

This scheme describes in a simplified way how individuals make decisions. Now, we can identify various points at which a BPP intervention might attack.

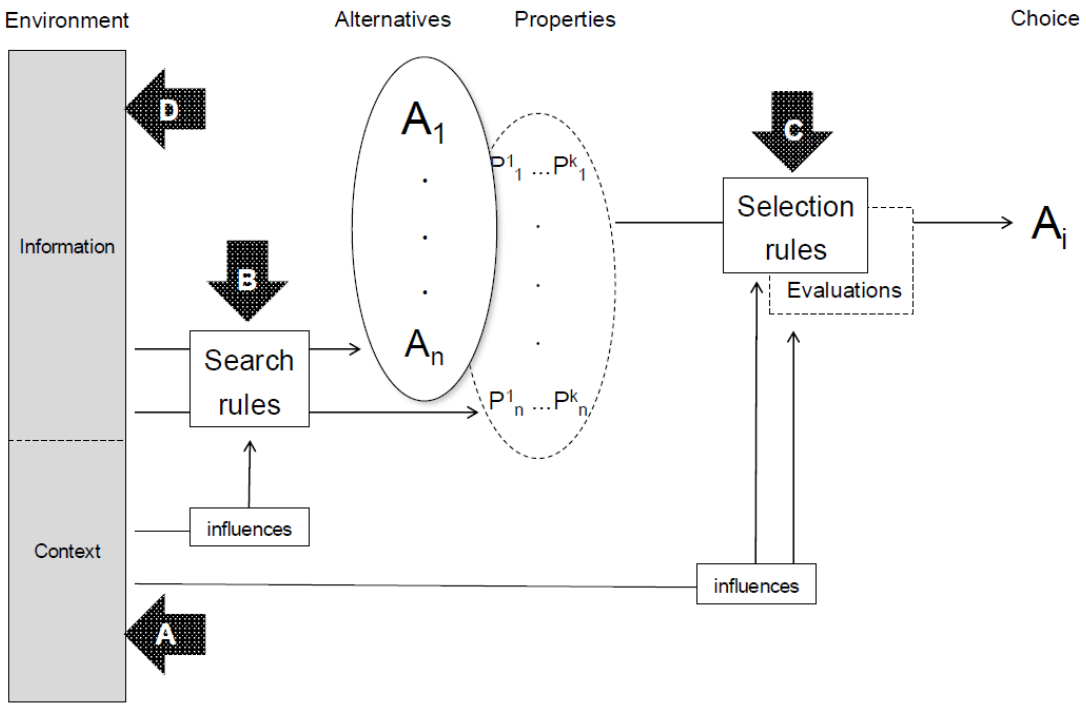

Figure 1. Distinguishing BPPs by their attack points.

With the help of this scheme, one can distinguish bias-harnessing from competence-enhancing BPPs. Bias-harnessing BPPs intervene on the contextual factors, removing, rearranging, or adding some, as indicated by arrow A in Figure 1. The intention is to exert influence through these contextual factors on the individuation and characterization of alternatives and on the search rule and the underlying evaluation. For example, the Save More Tomorrow intervention offers deliberators the choice between more consumption in

one year's time and higher pension payouts later, thus avoiding the "present bias" impact of a choice of more consumption now on the evaluation of the alternatives.[23]

Furthermore, bias-harnessing BPPs operate by harnessing factors that the targeted individual herself regards as irrelevant, but that have been shown to have an influence nevertheless. This corresponds to the pathways in the lower part of Figure 1. For example, individuals might consider irrelevant what option is set as a default, yet more people might end up with a certain option when it is the default, perhaps because they find the deliberation process too cumbersome and therefore never make an active choice.

Admittedly, sometimes the same intervention lever can operate through different mechanisms. Default-setting policy interventions, for example, have been speculated to operate either though cognition cost avoidance, status-quo bias, or receiving recommendation signals.[24] But in those cases, in order to determine *where* the intervention lever comes in, one must refer to the mechanism: a loss-aversion-driven default-setting policy, for example, comes in through an intervention on context factors, while a recommendation-driven default-setting policy operates through enlarging the searchable information set and thus comes in through an intervention on relevant information.

In contrast, competence-enhancing BPPs have a different entry point and operate through different mechanisms. They intervene on the search and selection rules directly, by teaching people new skills, training existing ones (arrows B and C), or by changing the environment to formats more fitting to acquired skills (arrow D). For example, training people on a new selection rule, such as "assume the perspective of a third-party spectator," to reduce marital strife[25] is an example of a C-intervention. In contrast, teaching people with little formal education a better search rule for their business purposes, such as "physically separate private and business receipts," is an example of a B-intervention.[26] To train people in transforming opaque representations (for example, single-event probabilities) into transparent

ones (for example, frequency-based representations) is a D-intervention.[27] Food recipes can enrich both search and selection rules (B and C). One can use them to search for what's available in one's refrigerator or at the market, or to select a dish for specific purposes, such as for a romantic dinner or for getting the right nutrition.

Elsewhere, I and Ralph Hertwig named competence-enhancing BPPs "boosts," suggesting that many so-called nudges fall into the bias-harnessing category.[28] Because there is no commonly agreed-upon definition of nudges, this claim has generated a lot of unnecessary controversy. What I claim here, instead, is that many prominent BPPs, such as SMT, as well as many framing and default-setting interventions are bias-harnessing. This claim suffices for my present argument.

We can now match the conditions identified above with this mechanistic distinction. First, because bias-harnessing BPPs intervene on effective contextual factors that individuals consider irrelevant, they might leave individuals in the dark about their own reasons for acting. They thus might not satisfy the transparency condition for autonomy. Second, because they operate through causal pathways separate from the individual's rational reasoning abilities, they might reduce their ability to self-reflect and deliberate about their choice. They thus might not satisfy the procedural-independence condition for autonomy. And third, because they have an influence on choice outside of individuals' reasoning capacities, they could easily contribute to individuals choosing against their authentic desires. They thus might not satisfy the non-alienation condition for autonomy. Note that bias-harnessing BPPs do not necessarily violate these conditions; some might be made transparent, some might not reduce self-reflection, and some might not alienate. But their potential to do so is clearly there, and deeply rooted in the mechanisms through which they operate. Consequently, bias-harnessing BPPs *might* be autonomy-reducing (and often are).[29]

Boosts (competence-enhancing BPPs), in contrast, seek to ameliorate search and selection rules, or improve their match with the environment. Yet such rules are part of agents' deliberation system, and their change requires the cooperation of targeted individuals. To be boosted successfully, subjects must at least be motivated at three stages: to expose themselves to the intervention, to engage with its content to acquire the competence, and to employ the thus-acquired competence for particular purposes. For example, if I tell you a new recipe, you need to take in what I tell you; when it's time to prepare dinner, you need to choose it over other recipes; and then you need to follow its instructions. Without such cooperation—that is, without the subject acknowledging that she has reasons to accept the boost—the boost will necessarily be ineffective.

Furthermore, the effectiveness of improved search or selection rules depends on the subject's using them in their reasoning and deliberation. Hence, an effective boost cannot lead to a reduction of subjects' ability to self-reflect and deliberate about their choices. Finally, because boosts operate through practical reasoning, they will depend on people's prior motivations and values. Boosts will not increase the alienation of the choices resulting from deliberations based on these. Consequently, due to their mode of operation, boosts *cannot* be autonomy-reducing.[30]

Enhancing agency

So far, I have shown that bias-harnessing BPPs might reduce autonomy, while boosts cannot. While this is a substantial result, it is cold comfort to those who understand that many people's autonomy is woefully low in many choice domains. Why prefer boosts to bias-harnessing BPPs for their better respect for autonomy, when there is lamentably little to be respected and preserved?

This is not a theoretical point. Many, if not most, people live their daily lives in "toxic choice environments,"[31] with choice architecture, incentives, and peer pressure aligned to serve private or public interests that are not their own. Examples include the food industry's profit-maximizing advertising, packaging, and production strategies with scant regard for health effects[32]; the pharmaceutical industry's tactic to report treatment benefits in the form of relative risks and harmful side effects in the form of absolute risks[33]; and the use of bias-harnessing BPPs by government agencies captured by lobbyists and vested interests.[34] Why would one care to preserve the autonomy of people in such circumstances? Toxic environments like these might already have reduced their autonomy to an extent that little might be preservable. Consequently, implementing a bias-harnessing, autonomy-reducing BPP that promises some other benefit would perhaps seem unproblematic.[35]

In such situations, the normatively more relevant question is whether a BPP is autonomy-promoting, not whether it is autonomy-preserving. This is what I aim to show for boosts: they can promote autonomy, specifically, by boosting agency. To argue this point, let's focus on agency.

Intentionality, in the philosophy of action, is widely seen as the mark of genuine agency. To act intentionally is explicated as acting for a reason.[36] This yields the distinction between intentional action (behavior performed for a reason) and non-intentional mere behavior (caused by something that isn't a reason). In Donald Davidson's account, an action is a behavior causally explained by a reason.[37] To illustrate, switching on the light to see what made that strange noise is an intentional action because the behavior is caused by the agent's reason to better see. If the agent instead brushed against the switch, inadvertently switching on the light, she would not have acted for a reason; her behavior would not be intentional, and we might call it "mere behavior."

To have a reason for action is, here, understood normatively: such reasons justify or make it right for someone to act in a certain way. Typically, this means that the behavior can be rationalized by a sound practical syllogism, consisting of the agent's goals and their take on how to attain them.[38] For example, detecting the source of the strange noise is a reason to switch on the light; flinching when one inadvertently turns on the light was brought on by a causal chain that did not involve practical reasoning at all.

Reflexes and impulses are examples of behavior that are neither intentional nor unintentional actions. Individuals who exhibit such behavior do not exercise agency. This distinction is important because bias-harnessing BPPs might trigger such agency-less behavior. To get clearer about this distinction, we draw on the "standard theory of action," a widely accepted account of what it is to act intentionally and for reasons. It says, roughly, that something is an intentional action and done for reasons just in case it is caused by mental states that the agent accepts as reasons, that is, that rationalize the action from the agent's point of view.[39] Such mental states in some cases might be desires and beliefs, but often will be intentions construed as irreducible mental states themselves.[40]

The standard theory, even in this rough presentation, provides a powerful tool for analyzing the relationship between BPP and agency. Human behavior sometimes lacks agency. This happens when reflexes or (nonrationalized) impulses cause behavior: those reflexes and impulses cause behavior without the agent accepting them as reasons for behaving in this way. But it also happens when subconscious or automatic causes, depending on contextual cues, are harnessed by third parties to influence behavior. As I mentioned above, nonconsciousness or automaticity does not exclude agency: agents can recruit the relevant routines automatically through conscious intentions and plans.[41] But this causal recruitment connection can be broken when third parties manage to harness these routines to

their ends, as bias-harnessing BPPs do. Such BPPs, thus, generate behavioral change while at the same time deteriorating agency.

The standard theory also offers the basis for understanding how agency can be improved. Agency is deficient when a subject makes a choice but does not accept the productive causes as her reasons for that choice. The purchase of a supersized meal, the acceptance of a risky treatment with little benefit, or the promulgation of disproven conspiracy claims, all made due to toxic choice environments, are examples of such deficient agency. Such deficiencies need not imply that agents have no reason at all for the choice. Having reasons for action, and hence acting intentionally, comes in degrees. Sometimes, one acts on a single reason, disregarding others that possibly speak against it. At other times, one carefully considers all reasons for and against an action before undertaking it or, to allow for automaticity, the adoption of habits was carefully reasoned for. Deficient agency thus consists in a low match between causes and reasons for a choice.

Inversely, increasing the degree to which that individual has reasons for that behavior enhances the individual's agency with respect to that behavior. This can be done either by rationally, noncoercively convincing the individual that the present cause is an acceptable reason for that behavior, or replacing the present cause with another one that is more acceptable as a reason for the individual. Thus, according to our interpretation of the standard theory of agency, enhancing agency involves tightening the connection between causes and reasons for choice.[42]

Enhancing agency matters because it is a pathway to promote autonomy. Above, I presented three necessary criteria for autonomy: having (and knowing) one's reasons for a choice, reasoning one's way to that choice, and the choice result not being alienating. But taken together, the three conditions do not constitute a sufficient criterion for autonomy. I might have a reason for a choice but not act on it; I might reason my way to a choice that

does not take my reasons into account, and my reasoning might not cause the choice; and I might choose a non-alienating result that is not caused by my reasoning and is not sensitive to my reasons. Each of these ruptures might be caused by a host of influences, including toxic choice environments and bias-harnessing BPPs. Consequently, a choice that meets the three minimal conditions is not necessarily an autonomous choice. There is more to autonomy than reasoning competence, reason responsiveness, and non-alienation. For the three criteria to be sufficient for autonomy, they must stand in the right productive relationship to each other.[43] Autonomy is established if non-alienating choice is *produced* by reasoning that is sensitive to the subject's reasons. This productive relationship is secured by agency, specifically, the match between causes of and reasons for action.

Boosts enhance agency

Again referring to the typology from above, I now argue that bias-harnessing BPPs cannot foster agency and hence cannot promote autonomy, while boosts can. As I already argued, bias-harnessing BPPs do not operate through reasoning; their causal pathways differ from rational deliberation. Furthermore, they either make the reasons for choice opaque for the subject or make subjects less responsive to these reasons. Finally, even if they improve choice with respect to alienation by making people choose more in accordance with their authentic desires, such BPPs do not produce this choice in an agency- and hence autonomy-promoting way.[44]

Contrast this with boosts. For boosts to be effective, they must co-opt the subject's motivation at the stage of acquiring, learning, and implementing the competence. In the first place, as I argued above, this prevents boosts from compromising the subject's autonomy. But it does more than that. By offering novel competences, or connecting extant competences with novel contexts, the intervention encourages subjects to examine their reasons for a

choice and engage in additional deliberation. Thus, it not only preserves but strengthens a subject's agency with respect to a particular choice domain. For example, when subjects ponder intervention to reduce marital strife, they need to consider their reasons for engaging in spousal arguments, for improving their relationship, and so on. They also need to deliberate whether this effort is worth the expected result. Similarly, when shopkeepers consider physical accounting intervention, they must consider their reasons for improving their accounting discipline and why the physical separation might be more suited for them than other actions, including doing nothing. The dorm residents who we provided with microwave recipes faced a choice between different energy-consuming cooking methods. For some of them, this was the first time they faced such a choice. The recipes thus challenged them to think about their choices and their reasons for them, in ways that they had not thought of before. Boosts thus not only preserve but positively nurture agency.

Their dependence on co-opting subjects' motivation and their fostering of agency ensures that boosts stand in the right productive relationship to the choice that they affect. Consequently, boosts can promote autonomy, if the thus-nurtured agency produces non-alienating choices.

Unfortunately, successful boosting does not guarantee that subjects employ the acquired competences only for non-alienating ends. One can easily imagine that a reluctant addict—that is, a person who craves to consume the addictive substance but does not endorse this craving—is more successful in satisfying her addiction due to acquired competences. More generally, one might think that a competence to achieve something constitutes a motivation to do it, even if upon reflection one would not value doing it. Consequently, boosting does not necessarily promote autonomy. Instead, successful boosting enhances agency in the boosted domain and, if the choices in this domain satisfy further conditions (chief among them that they yield non-alienating results), then boosts also promote

autonomy. Because boosts by themselves do not ensure that agents pursue only non-alienating ends, I conclude that boosts have the potential to promote autonomy, but only under the right conditions. The contrast with bias-exploiting BPPs, then, is that they cannot promote autonomy, while boosts can.

I want to finish with a speculation that, if true, would further strengthen my result. So far, I argued that boosts can promote autonomy, while bias-harnessing BPPs can't. Boosts don't always promote autonomy, because enhanced agency might be employed to further alienating ends. Agential competences might just as well enable people to do what they don't really desire or value (or even actively desire not to desire), as it enables them to pursue their genuine ends. Making the reluctant addict more competent in acquiring and self-administering the drug, for example, not only makes her worse off, it also decreases her autonomy, as it facilitates her satisfying her alienating addictive desire.

Allow me to push back against this view that boosts do not always promote autonomy because they do not necessarily promote authenticity. It seems to rest on the idea that people pick up competences in some value-neutral way: learning is neither affected by what values people have, nor does it affect these values or desires. But that strikes me as questionable.

First, as I argued above, competence acquisition requires people's cooperation. Boosting competences requires subjects to reflect on the reasons for adopting and implementing such competences. Through such reflections, subjects engage with their existing desires and impulses, and therefore are more likely to form or revise judgments about them. Hence, people's choice to acquire a competence constitutes an opportunity for self-reflection about the ends that these competences can further. For the addict, the question of whether to acquire new drug-procurement and -administration competences is exactly the moment to bring her reluctance to bear. This should give us hope that boosts tend to encourage agents to pursue only non-alienating ends.

Second, preferences, including second-order ones, are often constructed when control abilities identify new possibilities.[45] Boosting competences create such novel control abilities and thus allow people to identify new possible courses of action. Indirectly, boosts might enrich people's normative imagination, providing them with new reachable alternatives that demand to be evaluated. For example, many people expand their food preferences by reading new recipes, and here the mere imagination of preparing the food is often enough, without ever tasting it. Given that such construction is largely self-directed, as agents voluntarily accepted the boost, which inspired them to imagine new possibilities, there is hope that the preferences constructed in response are non-alienating.

Consequently, if boosts encourage self-reflection about existent ends and foster the construction of new preferences, which I admit is mere speculation at this point, then boosting competences might increase the chance that choices are driven by authentic and endorsed motivations, thus reducing alienating choice. In that case, boosts would not only have the potential to, but would actually, promote autonomy.

Conclusion

I offered a systematic comparison of two different types of BPP in terms of their impact on the personal autonomy of their targets. I proposed a categorization of BPP on the basis of the mechanisms through which they operate. Bias-harnessing BPPs intervene on the choice architecture and influence choice by using existing cognitive biases. I pointed to prominent nudges like SMT as well as various framing and default-setting interventions as examples of this type. Competence-enhancing BPPs (which I have termed "boosts") intervene on search and selection rules directly and seek to influence choice by making those rules correspond better to the relevant choice contexts.

I then compared these two BPP types with respect to three minimal criteria for autonomy preservation: reason-sensitivity, reasoning, and non-alienation. I showed that bias-harnessing BPPs might reduce autonomy, while boosts never do. I also argued that such autonomy-preserving conditions are of limited usefulness for BPP assessment, as many or perhaps most people live in contexts that substantially compromise their autonomy.

I therefore shifted the discussion to autonomy-promotion, making use of the standard theory of agency for this purpose. Agency ensures that the three minimal criteria stand in the right productive relationship to each other. This relationship, I argued, provides a sufficient condition for autonomy: Autonomy is established if non-alienating choice are *produced* by reasoning that is sensitive to the subject's reasons. Boosts, I argued, foster agency because they require that agents identify the reasons for adopting and implementing a new competence and because they force agents to engage in deliberation connecting these competences to their goals. Bias-harnessing BPPs, in contrast, cannot foster agency. Consequently, boosts can promote autonomous choice—namely, when the resulting choice is non-alienating—while bias-harnessing BPPs cannot.

These two results provide pro tanto reasons for policymakers to prefer boosts to bias-harnessing BPPs. This seems uncontroversial if boosts were at least as good as bias-harnessing BPPs in other respects, for example, effectiveness. But I believe that it also provides reasons to prefer boosts when they are not as effective as bias-harnessing BPPs, in cases where autonomy considerations matter. I do not, however, advocate a general preference for boosts based on these results: effectiveness and welfare considerations sometimes might trump autonomy considerations. Furthermore, there might be other BPPs besides boosts and bias-harnessing ones that might be preferable in some contexts. All this essay shows is that when considerations of autonomy reduction and in particular autonomy promotion matter, boosts have an advantage over bias-harnessing BPPs.

Acknowledgments. Financial support from the Swedish Energy Agency (Energimyndigheten, grant nr. P2022–01068) is gratefully acknowledged.

Competing interests. The author declares none.

[1] S. Lakshmi et al., "Energy Consumption in Microwave Cooking of Rice and Its Comparison with other Domestic Appliances," *Journal of Food Engineering* 78, no. 2 (2007): 715–22.

[2] Microwaving a 200g potato takes about seven minutes, while baking it takes about forty-five minutes. A microwave consumes about 1,000W, while an average electric oven more than 3,000W. Hence, energy use for a potato is about 117Wh in the microwave versus 2,250Wh in the oven.

[3] Those BPPs are often referred to as nudges. See Richard Thaler and Cass Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Penguin, 2009). However, there is no agreed-upon definition of nudges. For recent discussions, see, e.g., Philippe Mongin and Mikael Cozic, "Rethinking Nudge: Not One but Three Concepts," *Behavioural Public Policy* 2, no. 1 (2018): 107–24; Luca Congiu and Ivan Moscati, "A Review of Nudges: Definitions, Justifications, Effectiveness," *Journal of Economic Surveys* 36, no. 1 (2022): 188–213. As I am proposing a new, mechanism-based categorization of different BPP types, I will eschew the term 'nudge' here to avoid confusion.

[4] Yavor Paunov and Till Grüne-Yanoff, "Boosting vs. Nudging Sustainable Energy Consumption: A Long-Term Comparative Field Test in a Residential Context," *Behavioural Public Policy* 10, no. 1 (2026): 42–67.

[5] BPPs, as I characterize them, comprise many types of interventions beyond the dichotomy I am discussing here, e.g., information provision. Hence, although the two types I will be discussing are mutually exclusive, they are not jointly exhaustive.

[6] Luc Bovens, "The Ethics of *Nudge*," in *Preference Change: Approaches from Philosophy, Economics, and Psychology*, ed. Till Grüne-Yanoff and Sven Ove Hansson (Springer Science & Business Media, 2009), 207–19; Daniel Hausman and Brynn Welch, "Debate: To Nudge or Not to Nudge," *Journal of Political Philosophy* 18, no. 1 (2010): 123–36.

[7] For a recent overview, see Andreas Schmidt and Bart Engelen, "The Ethics of Nudging: An Overview," *Philosophy Compass* 15, no. 4 (2020): e12658.

[8] For reviews, see Adrien Barton and Till Grüne-Yanoff, "From Libertarian Paternalism to Nudging—and Beyond," *Review of Philosophy and Psychology* 6, no. 3 (2015): 341–59; Schmidt and Engelen, "The Ethics of Nudging."

[9] Hausman and Welch, "Debate: To Nudge or Not to Nudge," 128.

[10] Thaler and Sunstein, *Nudge*.

[11] Bovens, "The Ethics of *Nudge*."

[12] T. Martin Wilkinson, "Nudging and Manipulation," *Political Studies* 61, no. 2 (2013): 341–55.

[13] John Christman, "Defending Historical Autonomy: A Reply to Professor Mele," *Canadian Journal of Philosophy* 23, no. 2 (1993): 281–90; Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68, no. 1 (1971): 5–20.

[14] Bovens, "The Ethics of *Nudge*," 209.

[15] Richard Thaler and Shlomo Benartzi, "Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy* 112, no. S1 (2004): S164–87.

[16] John Christman, "Autonomy and Personal History," *Canadian Journal of Philosophy* 21, no. 1 (1991): 1–24; Christman, "Defending Historical Autonomy"; Alfred Mele, "History and Personal Autonomy," *Canadian Journal of Philosophy* 23, no. 2 (1993): 271–80; Alfred Mele, *Autonomous Agents: From Self-Control to Autonomy* (Oxford University Press, 1995).

[17] Bovens, "The Ethics of *Nudge*"; Pelle Guldborg Hansen and Andreas Maaloe Jespersen, "Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy," *European Journal of Risk Regulation* 4, no. 1 (2013): 3–28.

[18] John Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge University Press, 1998).

[19] Chris Mills, "The Heteronomy of Choice Architecture," *Review of Philosophy and Psychology* 6, no. 3 (2015): 495–509, 503. Cf. also Cass Sunstein, *The Ethics of Influence: Government in the Age of Behavioral Science* (Cambridge University Press, 2016), 54–56.

[20] Schmidt and Engelen, "The Ethics of *Nudging*," 5.

[21] Katherine Milkman et al., "Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling," *Management Science* 60, no. 2 (2014): 283–99.

[22] Till Grüne-Yanoff et al., "Toward a Framework for Selecting Behavioural Policies: How to Choose between Boosts and Nudges," *Economics & Philosophy* 34, no. 2 (2018): 243–66; Till Grüne-Yanoff, "Behavioral Public Policy: One Name, Many Types. A Mechanistic Perspective," in *The Routledge Handbook of Philosophy of Economics*, ed. Conrad Heilmann and Julian Reiss (Routledge, 2021),480–93.

[23] Thaler and Benartzi, "Save More Tomorrow™."

[24] Till Grüne-Yanoff, "Why Behavioural Policy Needs Mechanistic Evidence," *Economics & Philosophy* 32, no. 3 (2016): 463–83; Jon M. Jachimowicz et al., "When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects," *Behavioural Public Policy* 3, no. 2 (2019): 159–86.

[25] Eli J. Finkel et al., "A Brief Intervention to Promote Conflict Reappraisal Preserves Marital Quality Over Time," *Psychological Science* 24, no. 8 (2013): 1595–1601.

[26] Alejandro Drexler et al., "Keeping It Simple: Financial Literacy and Rules of Thumb," *American Economic Journal: Applied Economics* 6, no. 2 (2014): 1–31.

[27] Gerd Gigerenzer et al., "Helping Doctors and Patients Make Sense of Health Statistics," *Psychological Science in the Public Interest* 8, no. 2 (2007): 53–96.

[28] Till Grüne-Yanoff and Ralph Hertwig, "Nudge versus Boost: How Coherent Are Policy and Theory?," *Minds and Machines* 26, no. 1 (2016): 149–83; Ralph Hertwig and Till Grüne-Yanoff, "Nudging and Boosting: Steering or Empowering Good Decisions," *Perspectives on Psychological Science* 12, no. 6 (2017): 973–86.

[29] Another reason why bias-harnessing BPPs are sometimes not autonomy-reducing is that people exposed to them might have very little autonomy to start with. I will discuss this in more detail in the next section.

[30] This impossibility holds for the local-autonomy concept I am advocating here: boosts cannot reduce a person's autonomy with respect to the domain of the boosted choice. However, because attention and cognitive processing are scarce resources, boosting deliberation in one domain could perhaps diminish reflection in another domain, thus reducing autonomy there. It is thus not obvious that the impossibility holds for a global notion of autonomy. I thank Erik Angner and Malte Dold for this point.

[31] Ralph Hertwig, "When to Consider Boosting: Some Rules for Policy-Makers," *Behavioural Public Policy* 1, no. 2 (2017): 143–61.

[32] Marion Nestle, *Soda Politics: Taking on Big Soda (and Winning)* (Oxford University Press, 2015).

[33] Gigerenzer et al., "Helping Doctors and Patients Make Sense of Health Statistics."

[34] Riccardo Rebonato, *Taking Liberties: A Critical Examination of Libertarian Paternalism* (Palgrave Macmillan, 2012).

[35] Mills, "The Heteronomy of Choice Architecture"; Sunstein, *The Ethics of Influence*.

[36] G. E. M. Anscombe, *Intention* (Basil Blackwell, 1957)..

[37] Donald Davidson, "Actions, Reasons, and Causes," in Donald Davidson, *Essays on Actions and Events* (Clarendon Press, 1980), 3–20.

[38] Robert Audi, "Acting for Reasons," *The Philosophical Review* 95, no. 4 (1986): 511–46.

[39] Alvin I. Goldman, *A Theory of Human Action* (Prentice-Hall, 1970); Alfred Mele, *Motivation and Agency* (Oxford University Press, 2003).

[40] Michael E. Bratman, *Intention, Plans, and Practical Reason* (Harvard University Press, 1987).

[41] Markus Schlosser, "Dual-System Theory and the Role of Consciousness in Intentional Action," in *Free Will, Causality, and Neuroscience*, ed. Bernard Feltz et al. (Brill, 2019), 35–56), argues that an action is intentional either if the action is initiated and guided by a conscious intention or if it is derivatively intentional. Actions are derivatively intentional if they are the manifestation of a habit that either has been formed by performing actions of this type with conscious intent or has later been endorsed by the conscious formation of an intention. The problematic cases of BPPs harnessing habits both break the link to earlier intentional formation as well as prevent later intentional endorsement, hence preventing intentionality and therefore deteriorating agency. See also Randolph Clarke, "Skilled Activity and the Causal Theory of Action," *Philosophy and Phenomenological Research* 80, no. 3 (2010): 523–50.

[42] Sanchayan Banerjee et al., "It's Time We Put Agency into Behavioural Public Policy," *Behavioural Public Policy* 8, no. 4 (2024): 789–806.

[43] David Enoch, "How Nudging Upsets Autonomy," *The Journal of Philosophy* 121, no. 12 (2024): 657–85.

[44] To illustrate, consider the case of addition. If I'm addicted, I might not endorse the productive causes of my behavior (the chemical effects of nicotine on the brain, or whatever) as reasons, and there might be a low degree of match between causes of and reasons for action. A bias-harnessing BPP, for instance, a shock-image campaign or a social shaming intervention, might get me to act contrary to my addiction. So, this increases the match between causes and reasons for action. But it does not meet the right productive conditions: the non-alienating choice is not *produced* by reasoning that is sensitive to my reasons, but rather through the shock images or the social shame. Hence, this bias-harnessing BPP is not agency-enhancing. I thank an anonymous referee for the example.

[45] Craig McKenzie et al., "Constructed Preferences, Rationality, and Choice Architecture," *Review of Behavioral Economics* 5, nos. 3–4 (2018): 337–60.