Biases, Bumps, Nudges, Query lists, and Zero Tolerance Policies

Sheldon Wein
Saint Mary's University - Canada

Follow this and additional works at: https://scholar.uwindsor.ca/ossaarchive
Part of the Law Commons, and the Philosophy Commons

https://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentaries/100

This Paper is brought to you for free and open access by the Department of Philosophy at Scholarship at UWindsor. It has been accepted for inclusion in OSSA Conference Archive by an authorized conference organizer of Scholarship at UWindsor. For more information, please contact scholarship@uwindsor.ca.
Biases, Bumps, Nudges, Query lists, and Zero Tolerance Policies

SHELDON WEIN
Department of Philosophy and International Development Studies Program
Saint Mary’s University
Halifax, NS, B3H 3C3
Canada
sheldon.wein@gmail.com

Abstract: Zero tolerance policies are often mistakenly thought to be the best way to deal with pressing social problems. However, most arguments for zero tolerance policies are either based on inaccurate premises or they commit the zero tolerance fallacy. This paper explores ways that we might counteract the bias in favor of zero tolerance policies by adding a query list to the choice architecture.

Keywords: bias, bumps, choice architecture, nudges, query lists, trust, zero tolerance.

1. Introduction

Despite the longterm trends showing that humans are regularly treating each other more and more civilly, zero tolerance policies still hold some appeal. While the empirical data show that zero tolerance policies are rarely the best way to deal with an unwanted behavior, many people seem to think of them as the default plan when confronted with the issue of how to reduce or eliminate some behavior.¹ This paper explores a suggestion about how to alter the choice architecture for those who might consider zero tolerance policies in the hopes that zero tolerance policies will be used less frequently and only when more likely to be appropriate.

2. Zero tolerance policies

Zero tolerance policies—at least paradigm cases of zero tolerance policies—have the following characteristics:

1. There is full enforcement of the rule.
2. There is no prosecutorial discretion.
3. There is a strict interpretation of the rule.
4. There is strict liability for rule violations.
5. There is mandatory punishment (even for first instances of rule violation).
6. The mandatory punishment is (relatively) harsh.²

¹ And arguments, when such even exist, for such policies usually take the following form: Premise 1: We want to do as much as we can to keep people from x-ing. Premise 2: Nothing does more to prevent x-ing than having no tolerance for x-ing. Conclusion: Therefore, we should adopt the following zero tolerance policy: Absolutely no x-ing. Though formally valid, it is just an instance of the petitio principii fallacy because the second premise begs the question.
² These are derived from Wein (2014), though I have altered them somewhat to incorporate more recent thinking. I do not claim that these characteristics are each necessary and jointly sufficient (so this is not a definition of zero tolerance policies). Rather these characteristics are either part of the identity of zero tolerance policies or they are implied by the existence of such a policy. On this identity/implication distinction, see Aristotle, Topics (Book 1, especially Ch. 5). For an excellent contemporary discussion, see Shapiro (especially pp. 8 -18). For more on definition, see Wein (1980) and Wein (1983).

Below, I will refer to these as the six elements of zero tolerance policies. Here I will briefly say something about each of these elements.

2.1. Full enforcement of the rule

If we have zero tolerance for x, then whoever is charged with enforcing the rule against x is expected to report any instance of x that comes to their attention. Otherwise, there simply isn’t zero tolerance.

2.2. Lack of prosecutorial discretion

When there is a zero tolerance policy in place, then whoever decides whether someone who has allegedly violated the policy should be subject to the policy has no discretion to say that, in the particular instance, an apparent violation should not be pursued or reported.

2.3. Strict interpretation

When we have a zero tolerance policy against some activity, there is a strict (and implicitly fairly broad) interpretation of what constitutes that activity. If there is a zero tolerance policy attached to an activity, there is no room for those who interpret the rule to “bend or twist” the rule to exclude a certain instance or class of instances of apparent rule violation.

2.4. Strict liability

When there is zero tolerance for x-ing, someone who has x-ed cannot offer either a justification or an excuse for having x-ed. The only relevant defense one can offer when one is alleged to have violated the rule to which a zero tolerance policy is attached is that one did not, in fact, do the deed in question.

2.5. Mandatory punishment

Related to the above is the fact that whatever the punishment attached to violations is, it must be applied in all cases where someone has been found to have violated the rule. Compassionate or consequentialist reasons for withholding, deferring, or lessening the punishment simply have no role. If you are found to have violated the rule then the punishment must be applied regardless of whether its application will have any positive effects (or even if it is know it will have negative effects).

2.6. Relatively harsh punishment

Zero tolerance policies usually have harsher than normal punishments attached to them. This is because what motivates the adoption of a zero tolerance policy is the view that there is a serious problem that needs to be dealt with in a forceful manner. Harsh punishment is simply part of this.
3. Choice architecture, nudges, and bumps

3.1. The architecture of choice

How people are presented with options available for choice often affects which choice they are likely to make. Even when one does not do something coercive or rewarding to encourage the choice of particular options or move people away from certain options, one can affect the outcome just by altering what has come to be called the choice architecture. Sometimes simply putting “None of the above” in a survey dramatically affects how frequently any of the other options are chosen. Sometimes how an option is described—“rapid transit” versus “commuter train,” or “$100” versus “Only $100”—will affect how many people will select it. And, to use common examples, making one of the options the default or displaying an option more or less prominently affects how many people will choose that option.

3.2. Nudges

For purposes of this paper I will understand a nudge to be a way of using the choice architecture to influence people’s choices or behavior not by means that are coercive or incentive-based but instead by use of heuristics and biases.3 I coerce you into selecting A rather than B if I threaten to harm you in a substantial way should you choose B. And I create incentives for you choosing A over B if I tell you that A goes along with a financial reward but B does not or if I attach a tax or penalty to B. This account of nudges diverges from the standard accounts, in that some writers (including Thaler & Sunstein, 2012) take nudges to be influences that are either positively paternalistic or altruistic. That is to say, for something to be a nudge (on the view I am rejecting), it must either be in the actor’s interest to be nudged or, while not harming the actor, be in the interest of others. Thus, to take a common example, making the default option when one fills out a form applying for a driver’s license be that one is an organ donor is taken to be an altruistic nudge. It is a nudge because simply checking a box on the form indicating one does not want to be an organ donor takes minimal effort. It is an altruistic nudge because the benefit done by being an organ donor does not accrue to the donor but to the recipient. Having the default option when one starts employment be that one participate in a plan to save for one’s retirement (where one merely has to check a box if one does not wish to participate) is a paternalistic nudge because the beneficiary is the person being nudged. Placing healthy foods in prominent locations in cafeterias is both paternalistic (the customer in the cafeteria is more likely to choose the nutritious option and live a longer, happier life) and altruistic (because the rest of us will be less likely to have to contribute to the costs of the person’s health care or wait for our own health care while professionals attend to problems that person would not have had had she chosen more nutritious food at the cafeteria). Given these senses of the term, a self-interested nudge or a vicious nudge is simply a contradiction.

In this paper I use nudge in the broader sense. If, in deciding whether to put the fruit or the chocolate mousse in the more prominent place in the cafeteria display, I decide on the mousse because I make more money on it than on fruit, or because I have taken a bribe from a local cardiologist looking for more work, I am nudging you to choose mousse rather than fruit for dessert. Using the choice architecture to get you to choose something nutritious is a nudge,

3 For an alternative (and narrower) account, see Kumar (forthcoming).
and so is using that same architecture to get you to choose something bad for your health. Of course, if I do the former I am a better (though less affluent) person than if I do the latter.

3.3. Problems with nudges and heuristics

Nudges work. Many more people will decide to donate their organs after death if that is the default option than if the default option is not to donate. And people are more likely to buy candy bars if the treats are placed at eye level on a rack they are facing while in the grocery checkout line. But there are at least two problems with the use of nudges.

The first is rather minor: They are unstable. In some—but by no means all—cases, once people realize that they are being or have been nudged, the nudge loses its efficacy. But this is really just a practical problem, suggesting we should use unstable nudges only in particularly urgent circumstances in which we really have no other viable alternative or when the situation is unlikely to be repeated frequently enough for people to alter their behavior to nullify the effects of the nudge.

A more serious problem, or so it is alleged, is that nudges fail to completely respect autonomy. Of course, the use of nudges is not a big violation of autonomy. If the state gets you to agree to donate your organs for transplant after you die simply by making that the default option when you renew your drivers’ license, that comes much closer to respecting your autonomous decision about what to do with your body post-death than the government would if it decided that those who decide not to donate may no longer avail themselves of medical assistance when they need it. Nonetheless, those who worry about autonomy—who have what I think of as an autonomy fetish—think that deliberately using nudges and heuristics does infringe on full autonomy because such devices get individuals to make choices in ways they might not otherwise have made them and do so by relying on features of their selves over which they lack fully informed control. Surely there is something to this. If I am nudged to eat the fruit for dessert rather than the chocolate mousse simply by the way the two items are placed in the cafeteria, I am treated in a less respectful way than if I select the fruit because I have been informed of the relative health effects of the two options. As Victor Kumar (forthcoming) aptly puts it, “nudges do not straightforwardly infringe upon freedom of choice, but they nonetheless violate autonomy by circumventing rational agency. In a word, nudges are manipulative.”

3.4. Bumps

Nudges rely on biases or heuristics. One can nudge by relying on the bias for the status quo or by activating a heuristic. Nudges work, according to Thaler and Sunstein (2012), because our minds operate using two systems. One system—Type 1 reasoning—is intuitive, fast, and largely unavailable to conscious observation. The other system—Type 2 reasoning—is deliberative and slower than Type 1 reasoning, and we are regularly conscious of its operation. For many decisions—such as those to which we attach little importance, ones made under stress, and those which must be made quickly—we rely primarily on Type 1 reasoning. (This is familiar from Kahneman.) Nudges and heuristics exploit this fact about human reasoning to bend our decisions towards the outcome preferred by the choice architect. But, of course, things are not so simple. It seems that humans have the capacity to enhance the intuitive system through repeated and

---

4 The government of my home province, Nova Scotia, refuses to use this nudge even though it is aware of the benefits. It has not provided a reason for its intransigence on this matter.
careful use of the deliberative system. Thus we can learn skills through techniques that initially seem most like the work of the deliberative system and thereby supplement our intuitive reasoning. According to Stanley and Williamson (forthcoming), skills are “rationally learned knowledge contents.” Once acquired, they operate much like the intuitive system. Thus, bumps, as I use the term, are ways of employing the choice architecture to alter people’s behavior without either coercing or altering their incentives through appeals to rationally learned skills.

Bumps are therefore thought to be more respectful of individual autonomy. (The fact that humans are so good at acquiring skills, and that doing so is both appealing and natural, is likely due to the fact—for whatever reason—that we have evolved to overimitate much more than other hominoids.) Thus, bumps occupy a middle ground between nudges and rational argumentation. (Of course, bumps also lie between rational argumentation and outright coercion.) If we could develop a bump to get people to minimize the use of zero tolerance policies (or to use them only when appropriate), that would be more respectful of their autonomous decision-making than a nudge would be.

3.5. Autonomy and choice

Influencing what people choose to do can be threatening their autonomy. At the most extreme ends we have coercion, which always deprives people of their autonomy and which, if severe enough, completely destroys it. At the other end we have rational argumentation which completely respects people’s autonomy. In between are various other methods of affecting the choices people make and some of them impose upon individual autonomy.

Zero tolerance policies are ways of affecting the choices people make, and they are closer to the coercion end than to the rational argumentation end of the scale. Roughly, the scale goes something like this: coercion is the least respectful of autonomy, changing the outcome by either conjoining it with a cost or a benefit (as taxation and subsidies do) is next, then nudges, then bumps, then persuasion short of rational argumentation, and, finally, full rational argumentation (without being conjoined with any of the preceding). In the next section I will introduce the notion of a query list. Query lists, I suggest, fall into the penultimate category. They are better than either bumps or nudges (or if they are bumps—and these categories have yet to be fully sorted out—they are among those least threatening to autonomy) because they suggest to people that there may be rational arguments they had not yet considered.

---

5 For an overview of the recent literature on this see Kumar (forthcoming). For a full account of the nature of skills that I am using here see Stanley and Williamson (forthcoming).
6 I have taken the term “bumps” from Kumar. I do not follow his account, which confines bumps to those which are paternalistic or altruistic. I use the broader account for reasons similar to those given for using a broader definition of nudges that he and Thaler and Sunstein (2012) employ.
7 Furedi (2011) claims that nudges denigrate moral independence, erode our capacities to make judgments of value, devalue the private sphere, and are rather ineffective. Except for possibly the last consideration, none of his concerns apply to my proposal. Indeed, I would think the use of query lists would enhance independence, and increase our capacities for making value judgments, while being quite effective. (The private sphere is—for obvious reasons—not relevant here.)
8 On overimitation, see Shipton and Nielsen (2015), and Sterelny (2007).
9 Note that those who have an autonomy fetish and object to nudges rarely are bothered by sin taxes on cigarettes or alcohol, though these are a far more invasive interference with autonomy than most nudges.
4. Altering the choice architecture for those contemplating a zero tolerance policy

Essentially, a zero tolerance policy is a plan for dealing with an alleged social problem. And it is a plan that, by its very nature, displays two things: (a) the sense that the problem being confronted is a serious one that needs firm and perhaps radical measures to handle it adequately, and (b) that there is a substantial lack of trust in those who (to use Aquinas’s phrase) are charged with the care of the community to remedy the matter by more normal, less draconian, means. Bringing both the harsh and blunt nature of zero tolerance policies and the extent to which they inevitably presuppose a lack of trust to the forefront would serve to help those contemplating adopting zero tolerance policies to consider other alternatives. What is needed, then, is a device that would bring to the attention of those contemplating adopting a zero tolerance policy all six elements of such policies and what adopting this type of policy response reveals about the levels of trust among various officials.

Checklists might appear to be an option. They have been shown to be an effective way to ensure that actions follow one’s considered judgments about what to do. The best known cases are those used by airline pilots and by doctors in surgical operating theatres. Though there is some controversy about the range and extent of their effectiveness, it is generally recognized that in circumstances where there are well worked out and widely agreed upon procedures that need to be followed (often in a particular order) to accomplish an agreed upon end checklists can be helpful in ensuring that no essential step has been neglected. It is unclear to me whether checklists are better characterized as nudges or as bumps. But, however we think of them, it is clear that they do not constitute a serious infringement on autonomy.

In theory a checklist could be developed that would be of use to those who have decided after careful investigation and weighing of the evidence for and against to adopt a zero tolerance policy about some activity. Perhaps—and no one knows whether this is the case because the conceptual work has yet to be completed to enable to sorts of rigorous empirical testing that would be necessary to find out—it is the case the zero tolerance policies are such that their effectiveness depends on all six elements coming together in the right balance. In this sense zero tolerance policies may be stable property clusters, though we are still waiting to know just which properties (and in what balance) instantiate that cluster to make it stable.10 But the situation that faces us at present is not one where a checklist could serve a valuable function. All we know at this point is that some people in positions of authority have a bias towards adopting zero tolerance policies when they really have not fully thought out the reasons for doing so. What is needed is a device for getting them to consider the many factors involved and reflect on every aspect of the policy they are considering.

5. A query list

I suggest that a query list is a good way to ensure that policy makers and implementers avoid the zero tolerance fallacy. What I am calling a query list is just a list of questions or queries designed to ensure that all, or at least the most important, aspects of the contemplated decision have been given at least minimal consideration. In this sense is might be seen as a bump that encourages people who might be inclined to adopt zero tolerance policies to “zero in” on why they think

10 One might then think of zero tolerance policies as having something like the structure of homeostatic property clusters. For a good overview and an account of stable property clusters, see Slater (forthcoming).
they need such a policy and, more importantly, to help them realize that better alternatives likely are available.

Of course, the content of an appropriate query list will vary from situation to situation. (See the appendix for a draft of a query list for school principals thinking of adopting a zero tolerance policy for those who bring drugs onto a secondary school campus.) But in all situations, the query list for those contemplating adopting a zero tolerance policy will serve to ensure that prior to implementing such a policy they have given at least some thought to each of the six main elements of the proposed policy. This might help them consider the option of adopting a less-restrictive alternative which might be better suited to the task. Just getting people to go through the six elements, one by one, and focus on whether they can imagine alternative policies and the extent to which the zero tolerance policy they are considering is based on a lack of trust (and whether that lack of trust is warranted) will, I think, frequently get them to see that other less draconian measures may be more appropriate. The reasons for developing query lists which focus on issues of trust when a zero tolerance policy is being considered is, I suggest, because it is a good way to encourage thinking about viable alternative to zero tolerance policies. At least the first five, and perhaps all six, of the features of zero tolerance policies listed in the first section of this paper display a lack of trust by planners in those who implement the relevant policy. A query list which makes this lack of trust salient is, therefore, likely to cause re-thinking of the potential plan (at least among those who are capable of re-thinking).

6. Conclusion

A good way to reduce the inappropriate use of zero tolerance policies, while also respecting the autonomy and authority of those who might decide that in the particular circumstances they face the best policy to adopt is one that allows for zero tolerance, is to have those who consider such policies first to go through a query list to ensure that each of the identified elements is actually needed. It might seem odd to advocate methods which respect the autonomy of others when those others are contemplating denying third parties their autonomy. Yet the best way to induce those who are inclined to deny autonomy to others to change their ways may be not to show them some of their own medicine but instead to treat them as we would have them treat others.

Acknowledgements: I thank Victor Kumar for informing me of the recent work on bumps and Thea E. Smith for her extensive help.

References

APPENDIX

Sample query list (to be used by school principals prior to introducing a zero tolerance policy for drugs into their school)

1. Briefly explain what prompts you to be considering a zero tolerance policy.

2. State explicitly what the zero tolerance rule is.

3. Since zero tolerance policies do not permit any cases where apparent rule violations are not reported, state why you do not trust teachers and other school officials to use good judgment were they to have some discretion about when to bring apparent rule violations to your attention.

4. Since under a zero tolerance policy you would not have any discretion regarding whether to go ahead and penalize a student found to have violated the rule, explain why you, as it were, wish to tie yourself to the mast regarding potential future cases. Do you not trust your future self to use discretion wisely.

5. Briefly describe a couple of cases where you would regret that the policy was in effect. (That is, describe two cases—imaginary cases—where you think it would be appropriate to use discretion to not apply the rule—even though it had been violated—but where you would not be able to use this discretion under the new zero tolerance policy.) Consider how this would make you feel.

6. Explain the benefits of restricting yourself to a strict and broad interpretation of the rule you are proposing. Consider whether the case you outlined in 5 might benefit from a narrower rule.

7. Zero tolerance policies do not allow for those who violate the rule to offer an excuse of justification for their rule violation. Offer the best instance of an excuse you can think of and explain why you think it inadequate. Do the same for a justification.

8. Since zero tolerance policies require that all cases of rule violation receive a mandatory and harsh punishment, think carefully about cases like those you outlined in 5 above. Are you confident that you can live with yourself being responsible for harshly punishing those whose behavior—taken in isolation—would not warrant such severe punishment. Consider a slightly less severe punishment and explain why you think it would not do the job.