Limitations of behaviorally informed policy under social interaction

Timo Ehrig, Konstantinos Katsikopoulos, Jaakko Kuorikoski, Samuli Pöyhönen, and Shyam Sunder

Abstract

Nudge and boost are two competing approaches to applying the psychology of reasoning and decision making to improve policy. Proponents of both the approaches claim capacity to enhance social welfare through better individual decisions. We question the validity of this claim. First, individual rationality is neither sufficient nor necessary for improving collective outcomes. Second, collective outcomes of complex interactions among individuals are largely ignored by the focus of both nudge and boost on individual decisions. We suggest that the design of mechanisms and norms can sometimes lead to better collective outcomes than nudge and boost. More generally, we present conditions under which the three approaches enhance social welfare. Furthermore, we argue that to reliably improve collective outcomes that depend on the aggregation of many decisions, it is necessary to understand the interface between the psychology of reasoning and decision making on the one hand and economics and policy on the other.

Keywords: Behaviorally informed policy, institution design, nudge, boost, individual decisions, social welfare

JEL Codes: C90, D71, D78, L51, P21

Introduction: Social Science Tools for Improving Welfare

Since their conception as scientific disciplines, the behavioral and social sciences have been concerned with improving individual and social welfare (Peck, Peabody, and Richardson 1897). Development of these disciplines was assisted by sources of data made possible by the new bureaucratic institutions of the emerging nation states and the conviction that there are knowable laws of human behavior that can be used to enhance the functioning and welfare of society.
Recent decades have seen new proposals for using knowledge from psychology as input to policy. Much of this “behaviorally informed policy” has been aimed at improving personal decisions of individuals, implicitly assuming that more effective pursuit of individual goals also yields better consequences for society.

This approach has also been applied to improving the well-being of people in situations where they need to coordinate their behavior, for example, in jointly managing resources. Such problems cannot be reduced to the individual level, nor do they need to be considered on a par with statewide macroeconomic problems. In order to distinguish them from both individual decisions considered in isolation and large systems of interacting decisions that directly affect the macroeconomy, we refer to such problems as belonging to the *meso level*.

In this article we focus on interventions that do not impose new rules or regulations on behavior. Such soft interventions attempt to influence behavior by changing cognitive and affective aspects of the situation, people’s motivation, or their decision competencies. Such interventions try to alter choices without changing the opportunity sets of the decision makers.

Bond (2009) suggests that two major imports from the psychology of reasoning and decision making into policy can be distinguished. They are Richard Thaler and Cass Sunstein’s (2008) *nudge* and Gerd Gigerenzer’s (2007) approach, which Grüne-Yanoff and Hertwig (2015) called *boost*. They respectively draw on the two major research programs of the psychology of decision making (see also Kelman 2011): the “heuristics-and-biases” (Kahneman, Slovic, and Tversky 1982; Kahneman and Tversky 2000) and the “fast-and-frugal-heuristics” (Gigerenzer and Selten 2002; Gigerenzer, Hertwig, and Pachur 2011) programs.
In a nutshell, nudge holds that people systematically behave irrationally and have trouble learning how to make rational decisions; they need to be nudged toward better decisions. According to Rebonato (2012; but see Sunstein 2015), nudges employ people’s psychological biases (e.g., sticking to defaults) for improving their decisions (e.g., donating organs). Boost holds that people can make good decisions, even better than homo economicus, as long as they are educated to choose the appropriate rule of thumb for each situation or to improve other competencies for dealing with the relevant risks, probabilities, and statistics (Katsikopoulos 2014; Gigerenzer 2015). This is how we interpret nudge and boost in this article, although it should be noted that the definition of nudge is sometimes debated (Grüne-Yanoff and Hertwig 2015).

While nudge and boost are theoretically and practically very different, they are similar in the sense that they both attempt to improve collective outcomes via the improvement of individual decisions. The examples presented in Thaler and Sunstein (2008) refer to a mixture of individual and social problems but, in the end, offer solutions only for individual problems where the meaning of improvement seems to be clear, such as eating healthier or earning a higher return on financial investments. Gigerenzer (2007) suggested educational interventions for boosting a patient’s, doctor’s or policy maker’s understanding of the results of a medical test—which also seems to be a clear improvement—and the ability to decide accordingly. While these interventions could certainly scale up to the societal level (Gigerenzer and Muir Gray 2011), they start by pursuing individual improvement.

These examples are premised on social welfare being an increasing function of the welfare of the individuals involved. But in many important policy problems, this premise does not hold. For example, just the opposite is true in public goods dilemmas such as paying the fare for using public transportation. Although everyone benefits from
the availability of a public good, each agent is better off free riding on others’ contributions to the public good. The frequent presence of such collective action dilemmas in meso-level problems has not been addressed by the advocates of nudge or boost. Before such proposals can be seriously considered for policy, better analysis is called for, the support of politicians and media publicity notwithstanding.

Furthermore, these specific behavioral proposals are curiously disconnected from research on behavioral change based on social psychology (as found, for example, in the field of public health; Michie, van Stralen, and West 2011). This suggests that the prominence of the discussed behavioral change paradigms is partly a function of the ease with which they can be conceptually integrated with other policy-relevant social sciences. We argue that this is clearly the case with respect to nudge-polices and economics, and that the situation could be improved by building a common framework to assess all individual- and nonindividual-based ways of implementing soft interventions. The present article is a start of such an analysis.

We argue that nudge and boost cannot reliably lead to improved social welfare at the meso level because, in general, individual rationality is neither sufficient nor necessary for social welfare. We show how the design of mechanisms and social norms (for brevity, we refer to this as design hereafter) can succeed in addressing some problems not addressed by nudge and boost. We suggest conditions in which nudge, boost, and design improve social welfare in meso problems. We conclude by arguing that to reliably enhance social welfare when it depends on the aggregation of many decisions, it is necessary to understand and manage well the knowledge interface between the psychology of reasoning and decision making on the one hand and economics and policy on the other. We point out that doing so requires changing the prevailing mindset in behavioral modeling in economics.
The Link Between Individual Rationality and Social Welfare

It is well established that in a variety of economic and social phenomena, the rational pursuit of individual goals in the neoclassical economic sense does not yield improved collective outcomes. Three of many well-known examples are the tragedy of the commons (Hardin 1968), the prisoner’s dilemma (attributed to Flood and Dresher by Poundstone 1992), and Arrow’s paradox (Arrow 1951). One hardly needs additional evidence to reject such individual rationality as a sufficient condition for improved social welfare. In the following subsection, we give a few simple examples that counter the sufficiency proposition. We then discuss the insufficiency of individual rationality for social welfare in the more complicated case of modern markets. Finally, we discuss examples of markets in which not even individual rationality is necessary to attain collectively desirable outcomes. Our goal is to suggest that, in behaviorally informed social interventions as elsewhere, one should be careful to avoid being too quick to infer social welfare from individual rationality.

Simple Examples

In public transportation, traveling without buying a ticket (in systems where compliance is monitored sporadically) saves money for the scofflaws but undermines the economic viability of the service. Low fines combined with a low probability of apprehension render ticketless travel individually advantageous. For example, in Berlin a single ride at the time of writing costs 2.60 EUR and the fine for ticketless travel is 40.00 EUR, yielding a breakeven point of apprehension once every 16 rides. Furthermore, enforcement is concentrated in the first few days of each month and the undercover inspectors are not that hard to detect. Thus, for each single individual,
the rational decision—according to the expected value criterion—is never to buy a
ticket. Yet, few would argue that Berlin, its citizens, or the global environment would be
better off without the U-Bahn. The social and individual welfares are not aligned.

Improving the rationality of individual decisions does not necessarily lead to a
desired collective outcome, unless the environment is implausibly populated by
individuals who include social welfare in their utility function. Nudging individuals to
devise better ways of detecting undercover inspectors would put some change in their
pockets but it would be a shortsighted gain leading ultimately to higher costs,
congestion, discomfort, and travel time. Furthermore, it is not clear what would be the
aim of a nudge, other than to move people toward this kind of individual rationality
that, in this case, means traveling without tickets. Making people more rational could
also mean improving their understanding of the game-theoretic aspects of the problem,
which is an educational approach, more boost than nudge. However, it is not clear if
either approach will work if people’s motivation is to save money in the short term.
Many individuals may value the benefits that public transportation bestows on them
and others; but counting on the existence of such benevolent preferences is not a
reliable foundation for public policy.

Another way to nudge would be to make the probability of apprehension seem
higher than it actually is (this is contrary to the idea of boost because no competency is
improved). Of course, this means intentionally misleading the public, which some might
see as morally indefensible. Interestingly, this approach was taken by the Chicago
Transportation Authority to reduce speeding by creating the perceptual illusion that
drivers are going faster than they actually are and seems to be applauded by its
proponents (Thaler and Sunstein 2008, p. 41). The Berlin transportation authority
attempts to create a cognitive illusion by having posters saying “it is not worth it not to

pay” inside the trains, and the U.S. Internal Revenue Service publicizes stories of exemplary punishment for tax evasion during the week of the April 15 filing deadline.

The unresolved disconnect between individual and social welfare in the above example is obvious, but consider an example from resource economics that is less straightforward. Individual users of shared but slow-to-replenish resources, such as fisheries and forests, have an incentive to maximize their respective shares, depleting the resource to the eventual detriment of all. Indeed, virtually all kinds of fisheries as well as forests in large parts of the globe have been depleted rapidly over the past century (the United Nations Food and Agriculture Organization 2014).

If the cost of changing their line of work is high enough, it is individually rational for fishermen to keep fishing until the yield drops below their personal breakeven threshold. The feedback of falling yields through the years of overfishing rarely stops overfishing until it is too late. Each fisherman may hope that others will drop out first, or that the fish stocks will not run out before he or she retires.

So, how could a nudge or boost be implemented in such a case? It is not clear what the irrational behavior to be corrected is, or what decision competency is found wanting, except maybe people’s understanding of the game. The answers are harder to find than in the U-Bahn example because it does seem rational, or at least reasonable, to keep fishing and, paraphrasing L. J. Savage (1951), “cross the bridge of a barren lake, river or sea when we get there.”

In the two collective action dilemmas described above, there is a mismatch of individual and social welfare, even though the latter is relatively easy to define. Defining the social welfare function for a meso policy problem is often normatively and politically problematic. Consider the seemingly clear cut case of organ donations. The use of defaults to encourage organ donation has been cited as the archetype of
successful nudging, on the assumption that more donations are obviously better. A critic of nudge, however, might challenge this assumption. Note first that the organ donor and the recipient are linked by a long supply chain of individuals and businesses, consisting of providers of surgery at either end, medicines, equipment, hospitals, pathology, storage, matching, nursing, recuperation, physical therapy, lodging, and transportation. Most transplants are paid for by the government or private insurance companies who spread the considerable costs among larger population pools. In other words, whenever anybody in a pool gets a transplant, all members of the pool pay small amounts, depending on the size of the pool.

As the critics might point out, the complexity of the social consequences of transplants is often underestimated. The complications do not have to do only with the direct costs. Resources available for medical procedures are also limited. There is a good chance that governments and insurance companies must squeeze some other aspects of health care for money to pay for costly transplants, such as vaccination of children or education on hygiene and healthy lifestyles. If the money for transplants comes, for example, at the expense of promoting healthy diets, justifying organ transplants would call for a rather detailed economic, and moral, analysis.

Of course, there are some clearer cases: Few people would object to using organ transplants to cure sick children. However, to the extent that organ transplants go to chronically ill, mostly older people, the result is an aging population that is more expensive to maintain. One could ask about the social welfare consequences of keeping some number of older folks alive for a few extra months or years with the help of transplants. Is it fair to the younger generation? Who is to decide that nudging applicants to be organ donors is good for society? How should that decision be made?
Another example of an effective nudge is the increase in pensions saving achieved by simply changing the default employee option. Yet it is by no means the case that increased saving is always macroeconomically beneficial. Even if it is assumed that the individuals targeted would save more if they were able to rationally optimize their consumption over their lifetime, it still does not follow that the whole society would be better off. These examples show that even prima facie straightforward cases of soft interventions can give rise to difficult problems regarding the relationships between individual and collective welfare.

**Individual Rationality and Social Welfare in Markets**

A commonly used argument in the late 20th century was that market failures are a result of biased individual behavior, and if individuals would only be rational, markets would be efficient (Camerer 1992). This idea prevails in the literature on nudge and it implies that building architectures that enable market participants to choose according to their supposedly true, logically consistent preferences will improve efficiency (Johnson et al. 2013). Yet, recent theoretical work in economics suggests that in markets where disagreement and speculation play a role, individual rationality is not sufficient for social welfare.

Standard macroeconomic analysis draws a link from individual rationality to social welfare by using rational expectations equilibrium (REE) models (Radner 1979). REE models make strong assumptions such as that the mapping of any information to prices is common knowledge (Ben-Porath and Heifetz 2011). This would mean that a 75-year-old retiree in Michigan knows the precise implications of a new asset-trading class in an emerging market across the oceans. Under such strong assumptions, REE models imply that prices aggregate and summarize all available information (Lucas...
1972; Plott and Sunder 1988). That is, the price system functions as an invisible hand that efficiently allocates the resources.\(^1\)

Under more realistic assumptions, however, the link from individual rationality to social welfare, mediated by the market, is more complex. As Keynes (1936, Chapter 12) pointed out in his beauty contest metaphor, in the real world, agents often disagree and thus form beliefs about the beliefs of other agents. Higher-order-belief dynamics can lead to price indeterminacy in reality (Hirota and Sunder 2007; Hirota et al. 2015). For instance, I may buy a house today at $500,000, even if I value it at only $400,000 for my personal use, if I believe that I can sell it tomorrow for $600,000 to somebody else who values it even more. In other words, in the real world, disagreements lead to speculation. It has been shown analytically that given disagreement, speculation is individually rational (Ben-Porath and Heifetz 2011). But speculation can have a seriously negative impact on social welfare. For example, if there is disagreement about the value of assets, and financial innovations such as derivatives are introduced, then investors will speculate and portfolio risks can increase (Simsek 2013). This is because, given disagreement, derivatives create the illusion of being insured; even though bets are placed against each other, they do not balance out and instead increase the total risk.

**Individual Rationality May Not Be Necessary for Social Welfare**

The insufficiency of individual rationality for social welfare does not rule out the possibility that the pursuit of individually rational decisions through nudging or boosting may still be useful. It is possible that individual rationality is a necessary

---

\(^1\) But even when the strong assumptions of REE models are met, problems such as multiple equilibria may emerge (Cagan 1956; Marimon and Sunder 1993).
condition for social welfare. If a nudge or a boost increases the chance of individuals making rational decisions, this may also increase the chance that meso-level collective outcomes are improved. But, as we point out next, some markets are efficient even without individual rationality.

Derivations of equilibria in economic theory use, largely for analytical convenience, an idealized market form called the Walrasian auction.² It is assumed that all participants maximize their own profit or utility when responding to the auctioneer’s solicitations. This individual rationality yields social welfare in the form of Pareto-optimal competitive equilibria, according to the first fundamental theorem of welfare economics. Well before the formal derivation of the theorem, Adam Smith's (1776, Book 1, Chapter 2, paragraph 2) celebrated words highlighted this link: "It is not from the benevolence of the butcher, the brewer or the baker, that we expect our dinner, but from their regard to their own self-interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages." In other words, pursuit of self-interest by individuals helps furnish society members with the goods they want.

Adam Smith's link turned out not to be necessarily, for at least some markets. In their papers published in the same issue of the Journal of Political Economy in 1962, Gary S. Becker and Vernon L. Smith pointed to two complementary aspects of this nonnecessity. Becker showed that if consumers randomly chose an element of their

² This auction supposes that a neutral auctioneer calls out a price and asks all participants in the market to submit the quantities each of them would prefer to buy or sell at the announced price. Those whose personal demand price exceeds the announced price furnish their personal quantities above that price; those whose personal supply price is below the announced price submit their personal supply quantities. The auctioneer adds up the individual quantities submitted for purchases and for sales and compares the two sums. If the total sum demanded exceeds the total sum supplied, the auctioneer makes a small upward adjustment to the price and repeats the process; if the sum supplied exceeds the sum demanded, the price is adjusted downward for the next round of solicitations. The auctioneer repeats this process until a price at which the total quantities demanded and supplied are equal. This is the price at which all desired transactions can take place and are simultaneously executed.
consumption opportunity set, as defined by their resource constraints, the demand function is downward sloping. That is, rationality is not necessary for downward-sloping demand functions. Vernon Smith, on the other hand, populated his classroom markets with a dozen or so students motivated to earn money through trading. The students traded in a specific form of market organization called a double auction, modeled by Smith as a simplified approximation of stock exchanges. Surprisingly, the outcomes of these markets approximated competitive equilibria even though the markets were not organized as Walrasian auctions and did not fulfill most of the assumptions made in the formal derivations of competitive equilibria.

Some 30 years on, Gode and Sunder (1993) reported another result that serendipitously synthesized Becker’s random choice agents with Vernon Smith’s double auction. In this double auction, they replaced Adam Smith’s profit-motivated human traders with a simple computer algorithm: buyers bid uniformly distributed random numbers between zero and their value and sellers bid uniformly distributed random numbers between their cost and an arbitrary upper limit.

Surprisingly, the outcomes of this market, populated with “zero-intelligence” traders, also approximated the predictions of competitive equilibria derived from assumptions that did not hold in the market. The zero-intelligence traders of Gode and Sunder have no memory, do not learn, and certainly do not optimize. The only form of rationality that can be attributed to them is “don’t throw your money away by buying above your value or selling below cost.” This weak form of rationality is, in any case, imposed by markets in which people are not allowed to enter into commitments unless they have resources to settle them.

Taken together, the results of Becker (1962), Vernon Smith (1962), and Gode and Sunder (1993) establish that individual rationality is not a necessary condition for
market efficiency. These results suggest that significant aspects of the collective outcomes of markets may be driven by the structure of these institutions rather than self-love or rationality of individuals.

More broadly, a lesson to be drawn from the examples of this section is that the link between individual rationality and social welfare is very much determined by the structure of the problem at hand. One of our aims in this article is to induce conditions for the possible relationships between individual rationality and social welfare, as a function of what Herbert Simon (1955, 1956) called the environment. Only with such knowledge can we meaningfully talk about the success of nudge and boost. Before doing so, we discuss the design of institutions in the same example-based way in which we discussed nudge and boost.

**Design: Mechanisms and Norms**

If the relationship between individual rationality and social welfare is more complicated than pure sufficiency or necessity, it may be worthwhile to approach the problem of improving social welfare by looking at the social institutions itself.

**Mechanism Design**

Mechanism design, and market design more specifically, is a well-known class of interventions at the institutional level. In market design, insights from game theory and experimental economics can be used to devise the rules of markets in a way that leads to a good match between, say, employees and employers or kidney donors and recipients.

Consider, for example, the entry-level labor market of young doctors (i.e., residents) in the United States. Before the establishment of a clearinghouse mechanism
(in 1952, revised in 1995) for matching residents and hospitals, the strategic behavior of both doctors and hospitals had led to inefficient practices that none of the parties were happy with (Roth 1984, 2008). Hospitals competed for good residents, and despite the uncertainty about their competence, students were often hired several years before graduation. On the other hand, students had to make their acceptance decisions without the chance to compare competing offers.

The clearinghouse mechanism replaced the individual negotiations between residents and hospitals with both sides submitting a rank-ordered list of their preferences, after which a centralized algorithm would produce a matched list. The matching mechanism was received well by medical students and hospitals and led to high participation rates. As Roth (1984, 2003) has shown, the algorithm is stable in the sense that it never leaves a resident-hospital pair unmatched if the two would have mutually preferred to be matched together instead of being part of a different pair. However, as more women entered the medical profession, the need arose to match graduating married couples with one or two hospitals or programs within geographic proximity. As these problems eroded confidence in the matching mechanism, the system had to be redesigned to suit the changed circumstances (Roth 2008).

The original young doctor labor market unraveled because the market players responded strategically to the actions of others. The solution was to set up a voluntary institution with a set of rules that most of the participants recognized would leave everybody, themselves included, better off. More generally, good market design leads to markets that (i) have sufficient depth, meaning that the market has enough participants so that individual preferences can indeed be effectively paired; (ii) overcome congestion so that the participants have enough time and information to arrive at good decisions; and (iii) are safe in that it is reasonable for individuals to engage in the rule-governed
market transactions instead of face-to-face out-of-market transactions or resorting to strategically entering false preference orderings into the market algorithm.

In its attempt to get participants to reveal their true preferences, this kind of institutional design may appear to be similar to nudge. But it is fundamentally different. The designed rules of the game are systemic, apply to all, are targeted explicitly to collective outcomes, and are not specific to individual decisions.

Can market and social mechanism design address the problems involved in applying nudge and boost to the problems discussed above? At least in some cases, the answer is yes. For organ donations, Roth, Sönmez, and Ünver (2004) designed a matching market that is safe and has substantially increased the number of matches between willing donors and recipients in need. In the case of overfishing, formation of secondary markets for fishing quotas or seafaring rights could conceivably, if designed correctly, provide incentives for reducing the catch in such a way that those with the best possibilities for exiting the market would do so and those most dependent on the activity would not endure prohibitive personal and monetary costs.

The subway ticket problem seems to be solvable by design, as well. Making the probability of being caught high enough and raising the fine would eventually lead to a situation where free riding is no longer economically advantageous, as in Lisbon where the fine-to-ticket price ratio can be as high as 150. Also introducing controlled entry/exit gates in subway stations is a physical way of imposing a high cost to free riding. It is important to note, however, that these are not soft interventions, and that there are well-known problems with using explicit incentives to influence behavior. A major issue with incentive-based design is that the very introduction of incentives can change behavior. According to motivation crowding-out theory, explicit monetary
incentives crowd out intrinsic motivation based on adherence to civic duties and the like (Ostrom 2000).

**Norm Design**

Modification of social norms is another approach to intervention on collective action problems at the social level. As shown by Cialdini, Reno, and Kallgren (1990), at least two kinds of social norms can have a powerful influence on human behavior: People are motivated by both what others do (descriptive norms) as well as what is socially approved (injunctive norms).

In recent work spanning the fields of social psychology, developmental economics, and anthropology, Cristina Bicchieri and her collaborators have shown how coordinated manipulation of both kinds of social norms can be used to improve collective outcomes (Bicchieri 2006, 2015; Bicchieri and Mercier 2014). Their central premise is that instead of arising from stable preferences, much behavior is based on conditional preferences. That is, people participate in socially advantageous institutions at their own cost only if they believe that (i) others participate in the same behavior and (ii) others think that everyone should participate.

For example, consider the case of Muslim women wearing a veil (Bicchieri 2006, pp. 14–15). Even though a large part of the population might individually regard wearing a veil as burdensome, even oppressive, many Muslim women believe that they should be wearing one, because they believe that others expect them to wear a veil, prefer them to wear a veil, and might sanction them in some way if they did not wear one. A similar argument can be made for someone who prefers trousers to wearing blue jeans to college.
Such cases can be considered cases of belief traps or pluralistic ignorance, where people regard their own reasons for action differently from others’ and mistakenly assume that other people’s compliance with the norm results from their valuation of what the norm stands for. In the absence of transparent communication about the issue, high rates of compliance are taken as additional empirical evidence for the genuine intentions of others, and pluralistic ignorance functions as a self-fulfilling prophecy stabilizing the norm-based behavior. Moreover, it has been observed that despite not valuing the norm unconditionally, people are often willing to enforce it by sanctioning violations as they try to convince others, and possibly even themselves, of their genuine intentions (Centola, Willer, and Macy 2004).

In situations such as this, intervening to improve social welfare is harder than in the mechanism design case. Individual-level nudges or boosts are expected to be of limited value if they instill compliance to norms only for some isolated individual(s). Additionally, institutional changes or changes in legislation are unlikely to produce long-term change unless they align with the local social norms and are therefore seen to be legitimate (Bicchieri & Mercier 2014) However, interventions on pluralistic ignorance itself must solve a collective action problem: In difficult situations where strong expectations are in play (e.g., child marriage, female genital cutting, and avoiding breastfeeding; Bicchieri and Mercier 2014), educating or incentivizing people one by one is not likely to succeed. No one wants be the first person to behave differently and be subject to sanctions by the others still following the older norm. Because the behavior is rooted in conditional preferences, it seems that to replace one social norm with another requires that the entire population’s behavior must be altered collectively at once.
To deal with concrete instances of such problems, Bicchieri and her coauthors (ibid.) have suggested a group of interventions that can generally be called norm design. For example, assembling a whole community in the town-hall tradition of the New England region of the United States to deliberate on an issue and eventually elicit simultaneous public behavior-related pledges from all can influence people’s normative attitudes and at the same time create an empirically founded belief that the other members also share the same belief and act accordingly. As Bicchieri and her coauthors argued, achieving these goals at the same time can be sufficient for changing conditional preferences and can be understood as altering an old norm or as setting a new one.

One can imagine this approach being applied to the problem of overfishing. If some respected and successful fishermen can be persuaded to explain the need for reducing the catch, and personally commit to do so, the desired change in collective behavior may well occur.

In sum, the examples of this section suggest that sometimes soft interventions targeted at formal institutions or social norms might be more effective for improving social welfare than the individual-level approaches of nudge or boost. Of course, it should be noted that institution-level interventions may require detailed knowledge of the relevant institutions or norms, and such information is not always available to policy makers. Moreover, norm modification or norm setting requires that a normative community exist in the first place. For example, interactions between agents cannot be completely anonymous as they are in the subway ticket problem. It is questionable whether norm-based intervention strategies, such as signs proclaiming “if you do not pay, your neighbor will have to pay for you,” will noticeably improve social welfare.
In this section we present the conditions under which nudge, boost, and design can improve collective outcomes in meso-level problems, based on the examples discussed above. The conditions are not deduced from mathematical principles or generalized from systematic reviews of empirical studies. Nevertheless, they are presented as a potential springboard for more formal analyses and for better integration of a dispersed empirical literature. Table 1 organizes the (provisional) conditions by five aspects of meso-level problems.

In Table 1, we use the term “social welfare function” (SWF) to denote the mapping from individual to social welfare. Individual welfare may or may not be fully determined by individual preferences. We do not take any normative stance toward the possible redistributive properties of the SWF and simply ignore the problem of interpersonal comparisons of utility.

**Preferences of individuals.** The interventions discussed in Thaler and Sunstein (2008) mostly describe situations where the true preferences of the “nudgees” are relatively homogeneous and known to the “nudger.” This assumption seems reasonable for some problems, such as avoiding traffic accidents, but questionable for others, such as eating healthy. In the latter, the one-size-fits-all approach of nudge faces serious challenges. For example, the nudge might not respect everyone’s preferences, and, the consequences of nudging might be unpredictable. Rebonato (2012) discussed in detail just how daunting the task of nudgers must be, when they have to uncover every person’s “System II” for thinking rationally—which Thaler and Sunstein (2008) alluded to as the ultimate authority on true rational preferences—given the unobservability of preferences and malleability of human behavior.

*Table 1*

**Conditions Under Which Nudge, Boost and Design Can Improve Collective Outcomes**
<table>
<thead>
<tr>
<th>Aspect of meso problems</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nudge</td>
</tr>
<tr>
<td>Preferences of</td>
<td>Known and</td>
</tr>
<tr>
<td>individuals</td>
<td>homogeneous</td>
</tr>
<tr>
<td></td>
<td>preferences</td>
</tr>
<tr>
<td>Social welfare function</td>
<td>Known social welfare function</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Form of social</td>
<td>No collective action</td>
</tr>
<tr>
<td>welfare function</td>
<td>problems and no</td>
</tr>
<tr>
<td></td>
<td>frequency dependence</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Modularity of</td>
<td>Intended behavior is modular</td>
</tr>
<tr>
<td>target</td>
<td></td>
</tr>
<tr>
<td>Engagement of</td>
<td>Existence of a psychological bias</td>
</tr>
<tr>
<td>individuals</td>
<td>(according to Rebonato, 2012)</td>
</tr>
</tbody>
</table>

On the other hand, since boost is targeted at competencies, it does not require knowledge of preferences. Boost should just improve each person’s ability to achieve his or her respective goals, no matter what they are (although boosting the decision-making and planning competencies of criminals might not be such a good idea).

Finally, the designer of a mechanism needs to make sure that people with radically different goals from those advanced by the institution do not engage with that 20
institution. But this is not an unreasonable restriction. For example, why would people not committed to organ donation sign up for it? A more important restriction is that, for norm design, people's conditional preferences must be known.

**SWF.** Nudge, boost, and design, all have to take into account the exact way in which the fulfillment of individual preferences leads to collective outcomes. These approaches could also be implemented without knowing the SWF, but then an increase in social welfare cannot be guaranteed. This is a well-established principle of design, but a main message of the present article is that it also holds for nudge and boost. The proponents of nudge and boost must, therefore, specify the SWF for each meso-level problem to be addressed by the proposed intervention.

**Form of SWF.** In the meso problems we have been discussing—even in the simple ones such as ticketless travel and overfishing—it can be catastrophic to simply assume that the SWF is increasing in individual welfare, as nudge and boost implicitly do. Nudge and boost cannot handle collective action problems where things improve only if everyone, or at least most people, act simultaneously. Nor can they handle frequency-dependent situations where individual effects depend on whether, and how many, others are subject to the intervention. Well-designed markets can handle these problems but, as discussed above, the market needs to have depth and safety and to be free of congestion.

**Modularity of target.** Consider a fictitious intervention aimed at increasing the entrepreneurship of business students by influencing their attitudes toward risk but after which an increase in road accidents involving these students is observed. As with the drugs designed to selectively kill cancer cells in a tumor, the effectiveness of nudge also depends on its neutrality beyond the intended domain of the intervention. This can be guaranteed only if the intended behavior is suitably modular in the sense that its
underlying processes do not overlap with the processes underlying other behaviors in nontrivial ways. In principle, the same condition should also hold for the competency a boost intends to improve.

The modularity of boosts can vary dramatically: Some simple and fast decision trees (Gigerenzer et al. 2011) are applicable only to specific decision tasks and information sets, whereas improving risk and statistical literacy probably affects a wide range of behaviors. The corresponding challenge for design is to avoid the crowding out of intrinsic motivation, as in the erosion of civic virtues by explicit monetary or other incentives (Ostrom 2000).

**Engagement of individuals.** According to Rebonato (2012), since nudge employs people’s psychological biases (e.g., sticking to a default) to improve their decisions (e.g., donating organs), in many cases nudges are not even revealed to the agents. In this sense, nudge does not require the motivation to learn and improve one’s own lot. Boost, on the other hand, does require, perhaps unrealistically, that people are motivated to improve their competencies. Mechanism design proceeds on the assumption of self-interested agents, which probably is the case at least sometimes, and does not appear to require any additional motivations, whereas norm-based interventions require that a norm-seeking community exist in the first place.

The above analysis implies that, for some problems, neither nudge, nor boost, nor design would work. Should policy makers just let things be in such cases? Let us exclude from consideration situations such as humanitarian crises in which politicians and business leaders debate if it is anyone’s right to intervene, for fear of disrupting profitable business relationships with tyrants, or the frequent crises created by rickety fishing boats overloaded with desperate refugees seeking safe havens.
Instead, as an example, consider the simple problem of paying for subway tickets. In most cities of the world, passengers using public transportation include some proportion of nonpayers. They include the poor, indigent, desperate, forgetful, unemployed, ignorant, stingy, anarchist, and of course, the cheats. It is reasonable to say that their numbers are not reducible to zero at any reasonable cost and effort. Should one view the prevailing mix of nonpayers in any city as an evolutionary social equilibrium that society is comfortable living with? This equilibrium may result in letting the poor, indigent, desperate, forgetful, unemployed, and ignorant take a free ride without spending the extra effort to catch them and make them pay, a sort of social subsidy program. Anarchists and others who may not want to pay on ideological grounds are also left alone to have the satisfaction of engaging in their protest as long as it remains reasonably silent and small in scale. The additional cost of enforcing morality on the stingy and the cheats may not be worth the effort and may inconvenience the vast majority of paying passengers. This equilibrium has a little bit of everything in it, including norms, enforcement, cheating, and also some softness to accommodate the desperate. Seen in this light, all it means is that everyone else pays $1.11 instead of $1 to cover for the 10% nonpayers, and this is an acceptable price to pay for a decent level of social harmony, peace, and civilized cohabitation. If society decides to do this for subway tickets, where else does this work? How far can this approach be taken?

At least in some cases, letting behaviors be, or even embracing existing behaviors, can be a solution. Think of the solution to the problem in the online sharing of music analogous to ‘fare dodging’, that is, the illegal online sharing of .mp3 music files. While the music industry major labels first tried to change the online sharing behavior (Renner, 2004), the final solution to the problem was to invent new business models that embrace the resources generated by youngsters sharing music, in
particular their alive and creative attention and feeling of belonging. Business models like Spotify capitalize on these resources, e.g., by tailored advertisement.

Bringing this back to the discussion above, sometimes it can help to think of what is good and creative in how people already behave, instead of starting with the assumption that behavior needs to be changed.

Concluding Remarks

There is a lack of empirical and theoretical research on the conditions for effectiveness of the different kinds of soft interventions. The conditions presented in the previous section are little more than tentative observations – albeit ones based on central characteristics of the different kinds of interventions. A proper evaluation, let alone implementation, of nudge, boost, or design requires integration of knowledge from diverse fields, such as anthropology, cognitive and social psychology, sociology, politics, administration, and economics. However, the disciplinary structure of scientific research gets in the way of transfer of knowledge across disciplines (Daston 2010). The different and sometimes incompatible vocabularies, methods of theorizing, modeling, and handling data serve as barriers. In this concluding section, we briefly remark on the interface between the psychology of reasoning and decision making and the fields of economics and policy that undergirds the proposals for behaviorally informed policy intervention.

First, note that it is not surprising that nudge and boost more or less ignore system-level effects, since both are rooted in the experimental psychology of individual reasoning and decision making. Early neoclassical economists (Jevons 1871; Edgeworth 1881) thought that the understanding of economics could be grounded in hedonistic psychology, couched in formalisms borrowed from the physics of energy. The concept of
utility maximization was not only an interface between psychology and economics but also seen as a bridge between the individual and system levels within economics. Later, even though economics distanced itself from hedonistic psychology as its behavioral basis, the idea of utility maximization was retained. Today, macroeconomic models are routinely built around the construct of the representative agent, that is, the idea that an economic system makes decisions as if it were a single utility-maximizing individual.

Prospect theory (Kahneman and Tversky 1979), rooted in the hedonistic basis of decision making, is one of the conceptual foundations of nudge (Katsikopoulos 2014). For many researchers, it has tied economics back to psychology. It is still utility maximization, albeit with modified probabilities and the shapes of utility functions (Friedman et al. 2014). This conceptual compatibility of economics and prospect theory, and the heuristics-and-biases research program (Kahneman et al. 1982) more generally, may be an important reason why nudge has been so readily accepted by many economists. This convenient interface, however, also led to the assumption that correcting deviations from neoclassical rationality at the individual level will automatically lead to improved collective outcomes. The point of the present article is that this assumption can lead to serious errors in policy interventions in meso-level problems where aggregate outcomes follow from non-trivial interactions among individuals.

Boost has not yet caught on much in economics. This could well be because its conceptual basis, the fast-and-frugal-heuristics research program (Gigerenzer et al. 2011), does not offer anything resembling utility maximization that could be neatly plugged into the extant structure of economic theory. Instead, it is based on different concepts such as aspiration levels, ordinal comparisons, and lexicographic orders, originating in the work of Herbert Simon (1955, 1956). Even though studies have
shown that one fast-and-frugal heuristic—the priority heuristic—can predict major
laboratory phenomena in decisions under risk better than prospect theory, and the
concepts of loss aversion and probability weighting are not necessary for explaining
decision making (Katsikopoulos and Gigerenzer 2008), the faith of most behavioral
economists in prospect theory remains unshaken.

This status of prospect theory fits with the broader situation in behavioral
modeling in economics. Even after seventy years of intensive efforts, following Von
Neumann and Morgenstern’s challenge to give empirical content to expected utility
theory, the ability of the theory to predict out-of-sample observations remains meager
at best. Most economists continue to believe in, teach, and even insist on expected utility
theory (Friedman et al. 2014). Prestigious journals of economics regularly publish
proposals for adding new twists and free parameters to utility functions in the hope of
gaining some predictive power. To fit the data gathered from the same population in
different contexts, it has become necessary to attribute to them coefficients of risk
aversion that vary over a range of two orders of magnitude. Given the tight grip
expected utility has on researchers, it may take time to come to terms with its empirical
failure; in chemistry, Lavosier’s oxidation theory took almost a century to be accepted
over the prevailing theory of phlogiston to explain combustion.

Furthermore, being based on psychological decision research, both nudge and
boost are curiously separated from the vast empirical literature on behavioral change
techniques related to public health issues (Michie et al. 2011). Part of the solution could
be to follow the model of connecting policy with research on risk literacy and
communication (Gigerenzer and Muir Gray 2011). In general, even those economists
who do look to psychology for inspiration have borrowed very selectively and, for the
purposes of policy, inadequately (Tuckett et al. 2015). Markets and organizations are
systems with their own characteristics that extend well beyond the domains typically studied by psychologists (Gode and Sunder 1993). Serious dialogue among psychologists, organization scientists, and economists may help the field move forward. The psychology of decision making, behavioral economics, and management and organization science still lack much-needed mutual integration to provide advice on how groups of people reason, make decisions, and ultimately give rise to system-level outcomes. This dialogue needs to extend beyond the usual approach of plugging a model of the heuristics and biases program, such as prospect theory, into the maximization toolbox. Additionally, the interesting and promising work on group decision making and fast-and-frugal heuristics (Hertwig, Hoffrage, and the ABC Research Group 2013) seems as yet not to be developed enough to provide guidance on regulating systems such as markets and organizations. True, some economists and organizational scientists have been developing models of how individual behaviors based on heuristics scale up and lead to collective outcomes (Heinemann, Nagel, and Ockenfels 2004; Christensen and Knudsen 2010), but such models still refer to specific situations. Much more remains to be done and we hope that this article stimulates more effort in this direction.
References


