Soft Paternalism and Nudging - Critique of the Behavioral Foundations

Pasche, Markus

University of Jena, School of Economics and Business Administration

20 December 2014

Soft Paternalism and Nudging – Critique of the Behavioral Foundations

Markus Pasche

Friedrich Schiller University Jena
School of Economics and Business Administration
Carl-Zeiss-Straße 3, D-07743 Jena
markus.pasche@uni-jena.de

December 2014

Abstract

This brief note rises doubts on the argument that nudging will help people to behave more rational in terms of their own preferences. This justification of soft paternalism overlooks some methodological problems of expected utility theory which are one of the roots of behavioral economics.

Keywords: soft paternalism; nudging; behavioral economics, utility theory, rationality

JEL Classification: B4, D03, D04

1 Introduction: nudging people to behave "more rational"

Insights from Behavioral Economics can be used for a more effective design of economic policy. The justification of policy measures often relies on conventional arguments such as market failure (e.g. externalities), or the theory of merit goods. One example is the broad literature about nudging consumers towards more sustainable consumption patterns (Pasche 2014, Shogren and Taylor 2008, Venkatachalam 2008). Here, nudging is used as a new policy instrument, and it becomes more popular to ask behavioral economists and psychologists for policy advice (Madrian 2014, Amir and Lobel 2008).

Soft paternalism, however, is more than that. Here, it is argued that deficiencies in individual rationality, such like (time-) inconsistencies, intransitivities, and cognitive biases, prevent individuals from choosing the optimal alternative – optimal in terms of their own subjective preferences. Henceforth, economic behavior which is coordinated on markets will typically not lead to efficient outcomes. Thus, beside market failures like market power, externalities or information asymmetry problems, also deficiencies of individual rationality might constitute an additional justification for governmental intervention. It appears possible to "nudge" people to the "right" way in order to behave more consistent with their own interests but without limiting their freedom of choice (O'Donogue and Rabin 2003, Thaler and Sunstein 2008).

Obviously, this touches deep questions of individualism and liberalism and is therefore discussed very controversal although the idea is to keep the individual's full freedom of choice (see Sugden 2008, and recently Schnellenbach 2012, Kirchgässner 2014, Binder 2014, Binder and Lades 2015).

It is exactly this idea of enabling people to behave "more rational" in terms of their own preferences which is discussed critically from a methodological point of view in this brief note. It will not be drawn into question that insights from behavioral economics about typical patterns of boundedly rational belief formation and decision making could be useful for a more *effective* policy design[1]. And the arguments are only indirectly related to the field of political philosophy whether nudging might be consistent with liberalism or not. The latter debate is more about the legitimacy of paternalism. Instead, I will concentrate on the methodological concept of rationality and its narrow interpretation in the axiomatic utility theory and thus on the question whether it is possible and meaningful to "improve" individual rationality by nudging. A couple of literature contributions have recently discussed already methodological obstacles which should only briefly repeated here (see Binder 2014 for an extensive overview). What this brief note emphasizes is that (i) if we start from the rationality concept in utility theory there is no logical and empirical basis for deriving knowledge about the "true" preferences, (ii) it is doubtful whether utiity theory is an adequate normative basis for a behavioral economist, (iii) the perception of "anomalies" and "deficiencies" is misleading, (iv) paternalism in all forms falls behind a Public Choice perspective of policymaking and thus lacks itself a behavioral foundation of social mechanism design.

2 What do we know about the decision maker's preferences?

Inconsistent or non-optimal decision making means in the first place, that the axioms of expected utility theory (EUT) are not capable to describe the observed behavior. It lacks descriptive and thus explanatory power in the sense of a positive theory. However, EUT is still taken as the ultimative normative benchmark for rationality. Therefore, empirical derivations from EUT (so-called "anomalies" or "deficiencies") are thus seen as an expression of bounded rationality so that to some extent the decision maker violates his own preferences. Therefore, he should be happy about nudges in the "right" direction. Such nudges are based on observed regularities in the empirical deviations from EUT.

But if we accept that observed behavior is only boundedly rational, the idea of *revealed* prefences does not apply. The possibilities to draw conclusions from observed behavior to the underlying preferences are very limited. Thus, nudging has to rely on "typical" patterns of EUT deviations so that we can "guess" the underlying preferences. However, we know that these deviation patterns might depend on a lot of determinants such like experience, cognitive competence, environmental conditions, visceral factors etc. But then one cannot be sure that a nudge brings

[1] This does not neccessarily mean that policy *should* use nudging technologies to influence the behavior.

the decision maker closer to his (alleged) own idiosyncratic preferences. They are simply an unobservable part of the explanans of rational choice theory. How can we be sure that the nudges do not reflect the ideas of the policymaker what the indivdual "ought" to prefer rather than his true preferences? In case of e.g. the Ellsberg paradoxon or framing effects, there are always various solutions how the individual could behave in a consistent way. If he does not, however, to which solution he ought to be nudged? If we observe hyperbolic discounting then we can nudge the individual by e.g. stimulating precautionary savings or by incentivicing a reduction of procrastination. The resulting behavior should be time-consistent with an underlying constant discount rate r. And how large r "should" be so that the behavior is consistent with the alleged "true" preferences?

In a model of e.g. optimal sin taxes (O'Donogue and Rabin 2006) it seems to be possible to derive optimal nudges. Those models presuppose a "given" utility function. But since we cannot conclude from observed behavior to the underlying utility function (or preferences), the latter is simply unknown. And even worse, when the individual violates the underlying axioms, the existence and uniqueness (up to a positive-affine transformation) of such a function is not guaranteed. Henceforth, these models presume an omniscient institution which is able to express the "unbiased" preferences of the individual in a utility function. Then, of course, it is trivially possible to design an optimal nudge. The point is that such a ficticious utility function represents more the needs of the modeler who is eager to derive policy advices, but it is by no means clear whether it represents the individual's "true" preferences. Thus, Binder and Lades (2015) suggest not to choose utility theory but theories of subjective well-being as a starting point.

Let us consider the various attempts to restore the explanatory - or at least descriptive - power of utility theory by modifying or relaxing the underlying axioms. The most prominent example is the (Cumulative) Prospect Theory (Kahneman and Tversky 1979, Tversky and Wakker 1993). The maximization of such utility functions which do not have the expected utility property delivers a consistent description of the observed decision behavior in a wide range of cases. What appears as an anomaly within the EUT framework might now be in line with the predictions of the Prospect Theory (PT). What prevents us from interpreting PT as a consistent representation of the individual's preferences about uncertain outcomes? Accepting and acknowledging these specifities in choice behavior which appear as anomalies in light of the EUT, delivers us information about the underlying preferences and beliefs as characterized by PT – but not sufficient information about how this individual would decide *if* it would be perfectly rational.

Some behavioral economists argue that it is possible to construct "informed preferences" (see e.g. Thaler and Sunstein 2003) which reflect what the individual would choose if he is adequatly informed and would have unlimited cognitive abilities. But it is unclear when a choice can be identified as being "informed". I don't see neither a logical nor an empirical way to derive the EUT conforming preferences from boundedly rational choices. And as long as there are leeways of interpretation which are utilized by the policymaker, one should be very sceptical about welfare improvement of nudges.

3 Are there good reasons for behavioral "anomalies"?

A proper design of nudging instruments has to rely on stylized facts about typical decision "anomalies". But why do we observe *regular* behavioral patterns of "anomalies"? If such regular patterns have emerged as an outcome e.g. of evolutionary adaption or individual learning, how could it be that individuals adapted to something sub-optimal which violates their own preferences? Here, I subsume all behavioral approaches of other-regarding or social preferences (including things like fairness, reciprocity etc.) under the axiomatic utility approach since these approaches do not draw the rationality concept into question. If we would accept that people are not maximizing something but use heuristics, rules of thumb, being influenced by norms, peer groups, making information processing errors etc. as suggested by prominent behavioral economists, then they clearly do not behave *act rational*. But trying to explain these oberseved behavioral regularities – apart from maximization – in an economic meaningful way brings us to an interpretation of rationality as "*rule rationality*" (Aumann 2008).

People have learned to behave in a way which is optimal or at least sufficiently good not in every specific decision problem but in average in a large set of similar problems when knowledge is typically incomplete or vague, and information processing and decision errors might occur. These behavioral patterns (rules) are robust and successful and could be interpreted as an outcome of evolutionary adaption or learning. Their behavior should not be seen as a priori deficient as the terms "bounded" rationality and "anomalies" suggest (Berg and Gigerenzer 2007). This aspect is emphasized by the "fast and frugal heuristics" paradigm of cogntive psychology (Gigerenzer et al. 1999). Other authors like Heiner (1983, 1988) argue in a similar way that behavioral regularities could be explained as a smart control of the impact of information processing and decision errors in a complex environment. Also in experiments it turns out that deviations from EUT could improve the material outcome rather than reducing it (Berg et al. 2011). Woodford (2012) demonstrates how "biased" preferences as described by Prospect Theory can induce favorable outcomes in financial decisions. So there might be "good reasons" for the rule-governed decision behavior – which is a broader und perhaps more proper notion of rationality than the EUT axioms (Aumann 2008, Vanberg 2004). These arguments do not rule out that it might be possible to improve the decision quality for a single choice act. But they deny to interprete every departure from EUT as a deficiency which deserves to be healed.

In addition, Binder and Lades (2015) argue that nudges could disencourage critical reflection of decision behavior and learning, and could thus undermine the autonomy. There is also experimental evidence for unintended negative dynamic effects of nudges (e.g. de Haan and Linde 2012). Thus they claim that only nudges which foster the ability of the individual to critically analyse his own behavior and to strengthen it's autonomy should be considered.

4 Why should expected utility axioms be the ultimate normative benchmark?

Above, it was discussed already that one could choose other starting points for a consistent description of preferences than EUT. One could object that alternative approaches such like Prospect Theory have its merits as a descriptive model but that the "heart" of the *normative* rationality concept should be EUT as a benchmark. The deepest methodological argument against soft paternalism and nudging draws the axiomatic foundation of this underlying normative rationality concept into question. Among the various methodological objections against this axiomatic approach, one of the most challenging is the consistency requirement. "Why be consistent?", Robert Sugden asks in his 1985 article, and he denies that inconsistency is a rationale for paternalistic intervention (Sugden 2008). What if I have fun (utility) to behave in an "inconsistent" way? What if I know that I have unresolved cognitive problems with countervailing preferences, but I accept this fact? What if fundamental uncertainty (not simply ambiguity) prevents any probabilistic representation of my preferences? What if I consciously accept the possibility of a regret in the future when making "time-inconsistent" choices (the term, again, presumes already the acceptance of EUT axioms)?

In the old German fairy tale of *Hans im Glück* (*Hans in Luck*[2]), Hans receives a chunk of gold which he then sequentially exchanges with various other goods with "obviously" declining value. Every single deal appears him as very beneficial. At the end he possesses nothing, and he states that he now is the luckiest man all over the world. In terms of EUT his behavior is a chain of intransitive choices because the endpoint (having nothing) is strictly prefered to the starting point (having nothing) which was dominated by having the piece of gold. But is there any *logical* reason to deny that Hans' self-assessment of his final idiosyncratic welfare ("being the luckiest man all over the world") could be proven as *false* (see Schmidt 1995)? Hans' behavior is in the first place a problem for the economic modeler who insists on certain axioms, but the latter is not Hans' problem.

If non-acceptance of utility theory axioms is per se irrational and if the axioms are per se non-debatable, then the rationality concept is not much more than a conspiracy theory (see also Jones 1994). But there might be still "good reasons" to decide in this or that way apart from a EUT representation, and even in those cases, behavioral heuristics might work well in the long run. In such cases it is as a matter of principle not possible to nudge sombody to the "right" direction because in lack of a normative benchmark it is undefined what "right" should be.

If nudging is a policy implication derived from behavioral economics, then (this strand of) behavioral economics still uses EUT as the normative core of the explanans, it is not really an alternative to orthodox economics. One could state that it is not much more than a "neoclassical repair shop" (W. Güth). Thus, those economists who see behavioral economics as a broader or new paradigm, should have some problems to justify nudging and soft paternalism as a tool to make individuals

[2]The story is similar to the English tale *The Hedley Kow*. It is clear that the story can also be interpreted in various ways apart from the viewpoint of axiomatic utility theory.

behave more in accordance to their own preferences and thus emnhacing efficiency in the markets. This idea deeply roots in standard neoclassical economics.

5 The Public Choice perspective: who is nudging and why?

Policymakers are in principle not more rational than others. But the idea is that they should implememnt general rules and mechanisms instead of making discreationary nudges. Even under the advice of psychologists or behavioral economists the primary rationale for a policymaker is that the nudge induces behavior which is favorable for him or his political agenda. The term "paternalism" suggests, however, that a policymaker decides according to the idiosyncratic preferences of the people. This falls strongly behind a modern Public Choice perspective and thus behind a reasonable behavioral foundation of policymaking. As we have seen, there are some obstacles to the methodological possibility of a welfare improving nudge which gives a lot of discretionary leeways to the policymaker to define what the "right way" or "welfare" is.

Since nudging and other forms of soft paternalism are exploiting behavioral regularities apart from the rational choice model, they might work better the less the targeted individual is informed about that. This is, however, very problematic in a liberal society. And it is even more doubtful as the policymaker will follow his own goal function rather than being the benevolent "father" (lat. *pater*). Irrespective of his ideological motivation and his own behavioral "anomalies", the logic of political competition will force him towards opportunistic decisions since each political agenda requires a majority of votes in order to be implemented. Moreover, this process of policymaking is influenced by external lobbying and the interests of the bureaucracy, as it is known from Public Choice literature. Henceforth, the intransparanecy of the process bears a lot of risks when policymakers implement measures which are based on manipulation of the individuals. Justified by welfare arguments which can hardly be proven, nudging could be used for any political agenda or for rent-seeking activites of lobby groups. If any form of nudging – justified by conventional policy goals apart from rationality-enhancement – should be considered, then a maximum of transparancy about the aim, design, and behavioral background is required – even though this might reduce its effectiviness. Any form of nudging which aims at enhancing people's rationality, however, would not only require omnisciency and super-rationality of politicians (sic!), but is also in a sharp contrast with any reasonable behavioral foundation of policymaking.

6 Conclusion

I support the claim that insights from behavioral economics might be useful for a more *effective* policy design, e.g. in environmental policy. It is possible to design the "rules of the game" in a way that people behave more pro-social, more sustainable, strengthening reciprocity and promoting trust and trustworthy behavior etc.. As

long as these measures are not implemented in a hidden intransparent manner and do not undermine individual's autonomy, I don't see why such instruments should be less justifiable than other (conventional) instruments which are shaping the behavior as well, such like taxes. But it is doubtful to claim that others know better than the individual how to decide according to his *own* idiosyncratic preferences (the argument by v.Hayek and Buchanan, see Binder 2014). Based on the insights of behavioral economics and the methodological discussion of rational preferences, however, it is also doubtful to justify the exact opposite claim which is the credo of some liberal economists – liberalism is *not* grounded on superior rationality of the individual. There is simply no sufficient methodological basis for *rationality*-enhancing nudges. What the government might eventually do, however, is to strengthen the individual's autonomy and his ability to critically reflect his behavior (Binder and Lades 2015). But it is the question whether this should be called libertarian (or soft) paternalism since this term is now occupied by different meanings. Furthermore, it is doubtful whether EUT should be the ultimate normative benchmark for the policymaker. These doubts are not grounded on a liberal philosophy, they are grounded in the methodological criticism of the underlying concept of preferences and rationality.

References

Amir, O., Lobel, O. (2008), Stumble, predict, nudge: How behavioral economics informs law and policy. *Columbia Law Review*, 108(8), 2098-2137.

Aumann, R. J. (2008), Rule-Rationality versus Act-Rationality. Discussion Paper Series dp497, Center for Rationality and Interactive Decision Theory, Hebrew University, Jerusalem.

Berg, N., Eckel, C., Johnson, C. (2011), Inconsistency pays? Time-inconsistent subjects and EU violators earn more. Mimeo.

Berg, N., Gigerenzer, G. (2007), Psychology implies paternalism? Bounded rationality may reduce the rationale to regulate risk-taking. *Social Choice & Welfare* 28, 337-359.

Binder, M. (2014), Should evolutionary economists embrace libertarian paternalism? *Journal of Evolutionary Econonomics* 24, 515-539

Binder, M., Lades, L. K. (2015), Autonomy-enhancing Paternalism. Kyklos 68(1), 1-25 (to appear, related working paper: Levy Economics Institute, Working Paper No. 800.)

de Haan, T., Linde, J. (2012), Good Nudge Lullaby. Mimeo.

Gigerenzer, G., Todd, P. M. and The ABC Research Group (1999), *Simple Heuristics that Make us Smart*. Oxford University Press, New York.

Heiner, R. (1983), The origin of predictable behavior. *American Economic Review* 73, 560-595.

Heiner, R. (1988), The necessity of imperfect decisions. *Journal of Economic Behavior and Organisation* 10, 29-55.

Jones, E. (1994), The Tyranny of A Priorism in Economic Thought. *History of Economics Review* 22, S. 24-69.

Kahnemann, D., Tversky, A. (1979), Prospect Theory: an analysis of decision under risk. *Econometrica* 47, 263-291.

Kirchgassner, G. (2014), Soft Paternalism, Merit Goods, and Normative Individualism. CESifo Working Paper Series: 4688

Madrian, B. C. (2014), Applying Insights from Behavioral Economics to Policy Design. *Annual Review of Economics* 6(1), 663-688.

O'Donogue, T., Rabin, M. (2003), Studying Optimal Paternalism, illustrated by a model of sin taxes. *American Economic Review* 93, 186-191.

O'Donogue, T., Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics* 90(10), 1825-1849.

Pasche, M. (2014), What Can be Learned from Behavioural Economics for Environmental Policy? Jena Economic Research Papers No.2013-020

Reimer, T., Hoffrage, U. (2006), The Ecological Rationality of Simple Group Heuristics: Effects of Group Member Strategies on Decision Accuracy. *Theory and Decision* 60(4), 403-438.

Schmidt, T. (1995), *Rationale Entscheidungstheorie und reale Personen: Eine kritische Einführung in die formale Theorie individueller Entscheidungen* . Marburg.

Schnellenbach, J. (2012), Nudges and Norms: On the Political Economy of Soft Paternalism. *European Journal of Political Economy* 28(2), 266-277.

Shogren, J. F., Taylor, L. O. (2008), On Behavioral-Environmental Economics. *Review of Environmental Economics and Policy* 2(1), 26-44.

Sugden, R. (1985), Why be consistent? A critical analysis of consistency requirements in choice theory. *Econometrica* 52, 167-183.

Sugden, R. (2008), Why Incoherent Preferences Do Not Justify Paternalism. *Constitutional Political Economy* 19, 226-248.

Thaler, R., Sunstein, C. R. (2003), Libertarian paternalism. *American Economic Review* 93(2), 175-179.

Thaler, R., Sunstein, C. R. (2008), *Nudge – Improving Decisions about Health, Wealth and Happiness*. Penguin.

Tversky, A. (1975), A Critique of Expected Utility Theory: Descriptive and Normative Considerations. *Erkenntnis* 9, S. 163-173.

Tversky, A., Wakker, P. (1993), An Axiomatization of Cumulative Prospect Theory. *Journal of Risk and Uncertainty* 7, S. 147-176.

van den Bergh, J. C. J. M., Ferrer-i-Carbonell, A., Munda, G. (2000), Alternative models of individual behaviour and implications for environmental policy. *Ecological Economics* 32, 43-61.

Vanberg, V. J. (2004), The Rationality Postulate in Economics: its Ambiguity, its Deficiency and its Evolutionary Alternative. *Journal of Economic Methodology* 11, 1-29.

Venkatachalam, L. (2008), Behavioral economics for environmental policy. *Ecological Economics* 67, 640-645.

Woodford, M. (2012), Prospect Theory as Efficient Perceptual Distortion. *American Economic Review* 102(3), 41-46.