Francesco Bogliacino, Cristiano Codagnone & Giuseppe Alessandro Veltri

THE BEHAVIOURAL TURN IN CONSUMER POLICY: PERSPECTIVES AND CLARIFICATIONS

By Francesco Bogliacino, Cristiano Codagnone & Giuseppe Alessandro Veltri

1. Introduction[4]

Since 2012, the European Commission (from now on EC) has started testing policy options resorting to behavioural experiments (van Bavel, Herrmann, Esposito & Proestakis, 2013). This is a further sign of policymakers' increasing interest in 'Behavioural Economics' (BE hereafter), with special regard to consumer protection policies. In fact, in the same vein, Cass Sunstein became Director of the Office of Information and Regulatory Affairs under President Obama, whereas Richard Thaler became an advisor to the UK BIT, or *Behavioural Insight Team* (Kahneman, 2013). The German Chancellery has been recruiting staff with profound knowledge in the area of psychology, anthropology and behavioural economics for the Unit of Policy Planning, Fundamental Questions and Special Issues. Most recently, some important policy papers have been presented to discuss the general guidelines for the implementation of this approach (Dolan, King & Vlaev, 2010; BIT, 2011; Blumenthal-Barby & Burroughs, 2012; Oullier & Sauneron, 2010).

BE is the branch of economics that studies deviation from standard assumptions of rational choice (i.e. bounded rationality), grounding instead the understanding of human behaviour on cognitive psychology; it essentially studies the mind as an information processing device, against alternative approaches more based on the metaphor of impulse response. By examining deviation from rational choice, scholars in this domain have been able to identify a series of systematic errors in judgments (biases) and dependence of choices upon frames.

---

[4] The main input of this article is our experience in the framework of EAHC/2011/CP/01, Framework Contract with reopening of competition - behavioral studies. We thank G. Gaskell, M. Porta, A. Chakravarty, E. Ciriolo, G. Grimalda, P. Ortoleva, R. Van Bavel for the many discussions and exchanges on this topic. Comments from one of the editors have been appreciated and helped us to improve the final version. The usual disclaimer applies.

The latter phenomenon helps explaining addictions, anomalous consumption, and other dynamically inconsistent behaviours (i.e. hyperbolic discounting of future events).

A main implication of BE is that it is possible to apply insights from bounded rationality theory to *correct* mistakes by consumers or to induce certain type of conducts in cases in which the behaviours is inconsistent, e.g. in absence of accomplishment of targets by consumers (insufficient retirement saving, failure to quit smoking, etc.). According to this framework, it is possible to nudge consumers, promoting the (properly and normatively defined) *right* choice. This paternalism is matched with a statement of defence of the freedom of choice or consumer sovereignty (Thaler & Sunstein, 2003; 2008). The apparent oxymoron of libertarian paternalism disappear once we realize the existence of framing effect in choice: it is argued that a simple modification of the choice architecture, without altering the set of options (i.e. without restraining the freedom to choose) is able to accomplish the result. This claim relies upon the idea that it is possible to counter-bias and de-bias consumers (Kahneman, 2011) exploiting the very same mental shortcuts that generate bias in judgment and framing effects.

An important asset by the 'Nudge movement' is the large use of randomized control trials (RCT from now on), which are systematically employed in BE. The use of methodologies mimicking the hard sciences (essentially the construction of reliable counterfactuals to estimate the impact of an intervention) is defended as means to provide accountability and implicitly to *filter out* value judgments from the design of policy options, in harmony with the *evidence based policy* mantra.

In this article, we present a critical appraisal of the arguments about applying nudges and a defence of this approach. In particular, our target are not the scientific results of BE, but their translations in the policy domain. In the discussion of the potential of the BE informed interventions we point out two main caveats which we think have not been stressed enough, one is methodological and the other is theoretical. At the methodological level, we would like to discuss the limits of causal claims in experimental methodology applied to behaviour. As we will explain, causality is essentially constrained to the theoretical domain and cannot free the policy intervention from debate on the context of interventions. The technocratic idea

that we can proceed by trial and errors testing every intervention, and in this way avoiding conflict is scientifically unsound.

Secondly, at a more *theoretical* level, the statement of value free interventions, based on choice architecture is also flawed. First of all, the very logic of BE is that a minimal criterion of intervention cannot be defined since it should be based on exogenous preferences by individuals, which are contradicted by the very logic of context dependent preferences at the core of BE.

In the remaining of the article, we first discuss these two points separately, then we present a taxonomy of interventions in the consumer protection domain, and then we briefly conclude.

2. Neoclassical Experimentalism

The emphasis of RCT on the stage of policy debate is not a novelty, but rather a return of an agenda proposed from the 1950s onward under the label of 'Classical Experimentalism'.

By all means the idea of policies informed by RCT has a strong appeal for the public opinion and in the decision-maker. Social sciences have been traditionally cursed by having to deal with ex post correlation and thus with the lack of robust basis for causal inference. Ex post correlations cannot solve the traditional issues of omitted variables, simultaneity and measurement errors, and as such in many cases are inconclusive. Technically, the question that scholars are always dealing with is if the correlation can be interpreted in a causal way, in the limited sense in which cause precedes effect, cause covariates with effect and alternative plausible stories can be excluded (Shadish, Cook & Campbell, 2002): using data where the generation process is not under control, i.e. the assignment of independent variables and measurement of dependent ones is not part of the design, the researcher cannot list all the alternative explanations and as such can only limitedly rule out threats to internal validity.

In the course of history, a large effort has been carried out to build simulation techniques or alternative modelling tools to predict the effect of policy interventions. However, these models are plagued with problem of indeterminacy. Even under very rigid assumptions, such as the traditional computational general equilibrium techniques of mainstream economics, they pose very limited restrictions on the aggregate predictions that can be made. As a result, these models are not falsifiable, can reproduce any specific pattern observed and can be simulated to compute the effect of an intervention, but cannot be validated (Debreu, 1974; Mantel, 1974; Sonnenschein, 1972).

As a result, the development of controlled experiments can only be seen as fresh air and valid instruments to inform the decision-maker. By balancing threats to validity across groups and by equating expected values on pre-test outcomes, controlled experiments are valid tools to address the problem of causally interpreting a correlation. By providing full replicability and "relatively" simple technical apparatus (if compared with computational general equilibrium techniques), they increase transparency. Nevertheless, there is a sort of immanent risk in evidence-based policy, which can be magnified by RCT. In other words, due to the logic of the political process, there is the risk that policy makers perceive the evidence as a tool to overcome critiques and speed up the approval of the intervention, instead of an instrument to rationalize the use of resources and increase efficiency. It is as if the use of indicators and evaluation protocols would reduce attrition and conflict. In other words RCT may become another weapon in the hands of 'technocrats' and contribute to increase 'democratic deficit'.

Instead, some caveats are in order. In fact, we should be conscious that there are limits in the kind of lessons that we can learn from experiments. Decision-makers should understand that causality is essentially constrained to the theoretical domain and cannot free the policy interventions from debate on their context of reference. The idea of black box estimation of causal effects through experiments allowing a conflict-free policymaking is impossible to accomplish.

On this issue a clarification is in order: methodologically one should make a clear-cut distinction between a causal description and a causal explanation. The former is the case in which we accomplish a general statement relating two items. The classical example is the relationship between flicking a light switch and turning on the light. The causal explanation

occurs when we are able to account for the fact that the relationship between the light switch and the light may fail in presence of a burned bulb (Shadish, Cook & Campbell, 2002). In this case we are able to isolate the pieces along the causal chain.

Another way to describe it is through the traditional distinction between the effect of causes and the cause of effects, meaning the distinction between identifying ex post causal impact of an intervention and identifying structural parameters that explain the drivers of a certain change of behaviour following an intervention (Heckman).

If we aim at intervening on the environment, we need to accomplish explanation and not just description. However, this target is complicated by two main problems. On the one hand, we need both a correct categorical description of the phenomenon and a correct matching between the items measured and the categories. Clearly we need good theory for this (non overlapping categories with high explanatory power) and good operationalization. On the other hand, we have to face the issue of contextual causation that often occur in the form of complex adaptive systems: in the social sciences, we are never able to isolate fundamental causes, and most of the time we are really searching for consequences of non redundant pieces of non necessary but sufficient contextual causes. As a result we always have problems to extrapolate to alternative settings, units, descriptively different treatments and observations (Cronbach, Ambron. Dornbusch, Hess, Hornik, Phillips, Walker & Weiner, 1980).

It is important to stress that the idea of policy intervention as experiments to be evaluated using counterfactual techniques is not a new one. It dates back to the end of the 1950s, when a *popperian* program of '*reforms as experiments*' was defined in the US as a response to, at the time apparently effective, Soviet planning approach. In this initial plan, the *classical* experimentalism was rigidly founded on the use of social indicators and a strong foundationalism with a claim on privileged knowledge based on methodological strategic choices (Campbell, 1969). This program was aimed at both maximum accountability and reduction of social conflicts, in a sort of *truth to power* approach.[5] At the same time, it

---

[5] The expression 'truth to power' was first used in 1979 by Wildavsky (1979), after which it has been often applied to evaluation and impact assessment as methods to ensure accountability of the way public money is spent. See Wildavsky, A. (1979). *Speaking Truth to Power: The Art and Craft of Policy Analysis*. Boston: Little, Brown.

accomplished the more secular task of justifying an intervention (or guarantee checks and balances) in face of any regulation, ultimately understood as the product of rent-seeking by stakeholders (Niskanen, 1971; Posner, 2001).

Historically, the rigid counterfactualism of classical experimentalism was formulated by Campbell and colleagues, (Campbell & Stanley, 1963; Shadish, Cook & Leviton, 1991) based on the epistemological statement that experimental and quasi-experimental research design is considered as the only possible way to recover the causal impact of intervention controlling for all possible confounders and covariates.

Concretely, that model remained dominant for only a decade and was never really fully implemented in the practice of policy evaluation. The history of impact evaluation of policy interventions has been largely based on more pragmatic accounts -and the more so at the European level if compared with the US, UK, and other international organisation (Martini, 2009) - although aimed at producing hard numbers (not necessarily sounds from a scientific perspective). Methodologically unsatisfactory, this approach is more appealing to the policy environments, because it leaves room for mediation among different interests in the choice of the indicators and increase bargaining power through the use of hard numbers as rhetorical arguments.

In the concrete case of the European Commission, an example of this distance from the 'Campbell approach' towards a negotiation-plus-indicators framework was present since the establishment of the Open Method of Coordination in 2000[6]; and impact assessment followed the general rule of awareness of the complexity of political dimension of policy design and evaluation.

Página10

---

[6] The EU Open Method of Coordination (OMC) was launched in 2000. It is the governance mechanism for the achievement of the EU Lisbon Strategy objectives, emphasizing evidence-based policy and the role of measurement indicators (Saltelli et al., 2011, p. 198). For each given policy domain and/or sub-domain, it involves defining a strategy for setting policy guidelines, gathering a set of indicators and periodically revising target and progress. Targets and indicators are defined through a series of meetings with the Member States. The OMC was introduced as an aspect of 'new, experimental governance', which is part of the response by the EU to regulatory shortcomings (Szyszczak, E., 2006). For references and discussions, see Saltelli, A., D'Hombres, B., Jesinghaus, J., Manca, A., Mascherini, M., Nardo, M., & Saisana, M. (2011) and Szyszczak, E., (2006).

At the methodological level, there has been a large debate in the social science over the Campbell approach, without emerging consensus. In the decades that followed its formulation, alternative paradigms emerged. Already in the 1970s, counterfactualism was challenged by the development of the utilization-focused evaluation (Patton, 1997). In the simplest way it can be considered as a version of philosophical pragmatism, resting on the notion that ideas become true just in so far as they help us to get into satisfactory relations with other parts of our experience. So, in utilization-focused evaluation, knowledge validity rests on usefulness to the political process rather than on any methodological choice justifying a claim on privileged evidence.

Another challenge based on philosophical stance was present in constructivist approach to program evaluation: in this case the focus stands on stakeholders, since the truth in this case is envisaged in the construction of meanings by the different actors (Pawson & Tilley, 1997).

Finally, in terms of concrete implementation, a more widespread approach to evaluation came in the *efficientism* of the 1990s (Visser, 2003). With the strict focus on value for money, it is a mix of hardening and softening of constraints: usually it relies on a mixed hybrid of hard and soft evidence aggregated quantitatively with best of breed tools (Codagnone, 2007). In other word, while the claim of cost-benefits analysis is at the core of this approach (i.e. being accountable for the use of public money), it ends up being a sort of pragmatism without disclaimer: results are aggregated without rigid methodology to discuss the recovering of causal impact, and similarly, user defined variables are mixed up without concrete justification to build indicators of the results.

In this framework, a return to counterfactualism took place at the end of the 1990s. In this case, it is mainly supported by the increasing interest by policymakers on BE. This can be dubbed a sort of Neoclassical Experimentalism, to distinguish it from the original program.

At the epistemological level, counterfactualism is grounded on a 'successionist' notion of causality derived from Hume (1739). In a nutshell, it suggests causality as a relationship between experiences and not facts. It requires temporal and spatial contiguity (of cause and effect), temporal succession (from cause to effect) and conjoint occurrence (of effect if cause is observed).

From a methodological point of view, an experiment is an ideal situation in which the data generating process of the variable of intervention is subject to exogenous variation (internal validity), the participants to the study are representative of the population object of the intervention (external validity), and finally the behaviour measured in the lab is (on average) identical to the variable of interests that we want to modify in the real environment (construct validity). In concrete, specific problem may emerge that distance RCTs from this ideal situation. In reality, experiments are very local type of evidence, while our aim especially in our interaction with the policy maker is to have a general conclusion.

Lack of external validity is an endemic problem: participation can only be voluntary and most of the times these studies rely on convenience samples. Construct validity itself is subject to a number of trades-off, namely between the control over the data generating process requiring simplicity of the tasks performed to avoid confounding factors that impact of the design, and the fidelity of the task to the aim of the intervention, which requires complexity of the behavioural variable registered. This also makes it clear that an experiment cannot free the social scientist (or the policymakers informed by it) from dealing with theory: in order to design the experiment we need to have some sort of model, i.e. a list of definitions that will guide the identification of the response variables (operationalization), a sketch of the intervening variables which we need to keep under control to avoid confounding factors (with measurement, assumptions or design) and some sort of assumption over behaviour (otherwise we will never be able to predict who will change behaviour once the policy is implemented).

The experimental evidence used in the policy domain usually comes from two different sources: RCT in labs and *natural experiments* (or field experiments) associated to the piloting of a phased intervention in which the policy is first tested on a part of the population and then implemented *erga omnes*.

It should first be said that the two pieces of evidence have very different features. In the lab there is more control of the environment and the concrete possibility to master or manipulate confounding factors - but in the end over some of the factors there must be some assumptions (Camerer, 2003) -, but the sample is usually self-selected or simply a convenience one. The problem of construct validity is strong. For example, there are situations in which the exact

behaviour cannot be replicated, and some proxy conduct is measured (e.g. the case of simulated purchase). Alternatively, one may aim to capture complex unobservable conducts such as trust or cooperation, and the experimenter should re-create simplified situations which are more demanding in terms of theoretical assumptions. Human behaviour that entails complex social interactions cannot be generated in a lab context in which at best small group interactions can be tested. Finally, the problem of experimenter demand effect is a serious threat. In many cases, participants rely on any sort of cues to understand what could be socially desirable and this may confound the results (Zizzo, 2010).

In the field, the construct and external validity are increased but the control of the environment is drastically reduced. Anything that may be going on during the experiment may have strong bandwidth effect (Mullainathan, Shafir, 2013) altering the cognitive resources used by the participants. In the case of phased intervention, obviously the set of confounding factors is potentially unlimited. Moreover, the internal validity can be seriously threatened by attrition or selection. Finally, contamination of untreated by treatment or other forms of spillovers may be possible.

In both cases, rarely experiments deal with medium run or long run effects, because of dropout of participants or too limited budget to conduct the study making a longitudinal follow up or recall impossible.

In a nutshell RCT are certainly transparent, and full replicability is a much valuable gain that other social science methods tend not to share. Having said that, the generalization of a tested policy is conditional on the equivalence between the implemented and the tested policy and on assumptions over how agents in different context respond to the tested intervention. There is not a blueprint for this stage, but for sure theoretical guidance is necessary, which clearly imply that we cannot easily overcome opposing opinions.

3. Value free interventions?

Traditional policy making (including laws and regulations) is heavily shaped by the perspective of the homo oeconomicus of standard economic theory (Barr, Mullainathan &

Shafir, 2013). Homo Oeconomicus can be defined as a subject equipped with a stable system of preferences and the cognitive resources to process the information, avoiding systematic mistakes. Behavioural economics and nudge depart from mainstream economic theory and its implications for policy-making.

The orthodox view of the economic agent is grounded on a mathematical precise formulation of rational choice theory (Von Neumann & Morgenstern, 1944). Under a certain set of assumptions, the choices can be represented as a maximization of utility (i.e. acting following a certain preference ordering and given the constraints). Rationality requires also some sort of consistency in the way evaluations of the probability of event is done and revised (namely, not violating Bayesian rules). Preference ordering is deemed exogenous and not influenced by the specific choice set and the logic is consequentialist, in that alternatives are ranked on the basis of outcomes. The policy implication for the *demand side* is represented by the information paradigm (Micklitz, Reisch & Hagen, 2011): if the above apparatus holds, giving more information must empower citizens.

Behavioural scientists have theoretically and empirically shaken this edifice and departed from the above axiomization showing that judgment relies on heuristics and choices are reference dependent (Camerer, 1995; Camerer & Loewenstein, 2003). Empirically, they have found that human behaviour is heavily context dependent, 'a function of both the person and the situation'.(Shadish, Cook & Campbell, 2002) There often is no given ordering of preferences at all, and the latter are rather constructed (Slovic, 1995); the framing of the situation affects the final choice (Tversky, Kahneman, 1974); and the ordering is affected by the endowment available at the timing of decision (Thaler, 1980), even at the point in which the ordering is reversed (Grether & Plott, 1979). Present bias of individuals pushes them to revise their planned choices when the temptation materializes, as patently shown by smoking and alcohol use (Loewenstein & Prelec, 1992).

Heuristics and dual process theories (Stanovich & West, 2000; Thompson, 2009) are key to understand both the behavioural critique to standard economics and the main thrust of the nudge approach from a consequentialist perspective.

On the normative side, the violation of standard axioms of rational choice and the presence of framing effects and context dependence implies the lack of invariance of preference

ordering when the set of constrains is modified. As a result, we cannot have a minimal criterion based on the preferences themselves to evaluate two different social allocations (the so called Pareto criterion). As an example, contrast between dual selves (as in the case of addiction) implies an intra-personal comparison between two different systems of preferences and a choice of one of the two is ultimately a value-loaded decision (Codagnone, Veltri, Lupianez-Villanueva & Bogliacino, 2014). Of course, we can still think at some sort of criterion based on pair wise coherence, where an option is preferred to an alternative if the latter is never chosen when the latter is available (Bernheim & Rangel, 2009), but this criterion will be mute precisely in these situations in which what we intend to do is not what we do, which is the main object of policy discussion.

Instead, all policy intervention has some sort of latent value judgment. Behavioural science is not dealing with that, but rather with another fundamental issue: every policy intervention has some sort of implicit assumptions over how consumer will respond. BE deals with these assumptions, eventually helping to revise them in order to better fit evidence.

As it has been argue, BE informed policy should be a delicate mix between normative choices to define the best options, descriptive account of the behaviour and prescriptive identification of the gap between the desired and actual outcome (Fischhoff, 2013).
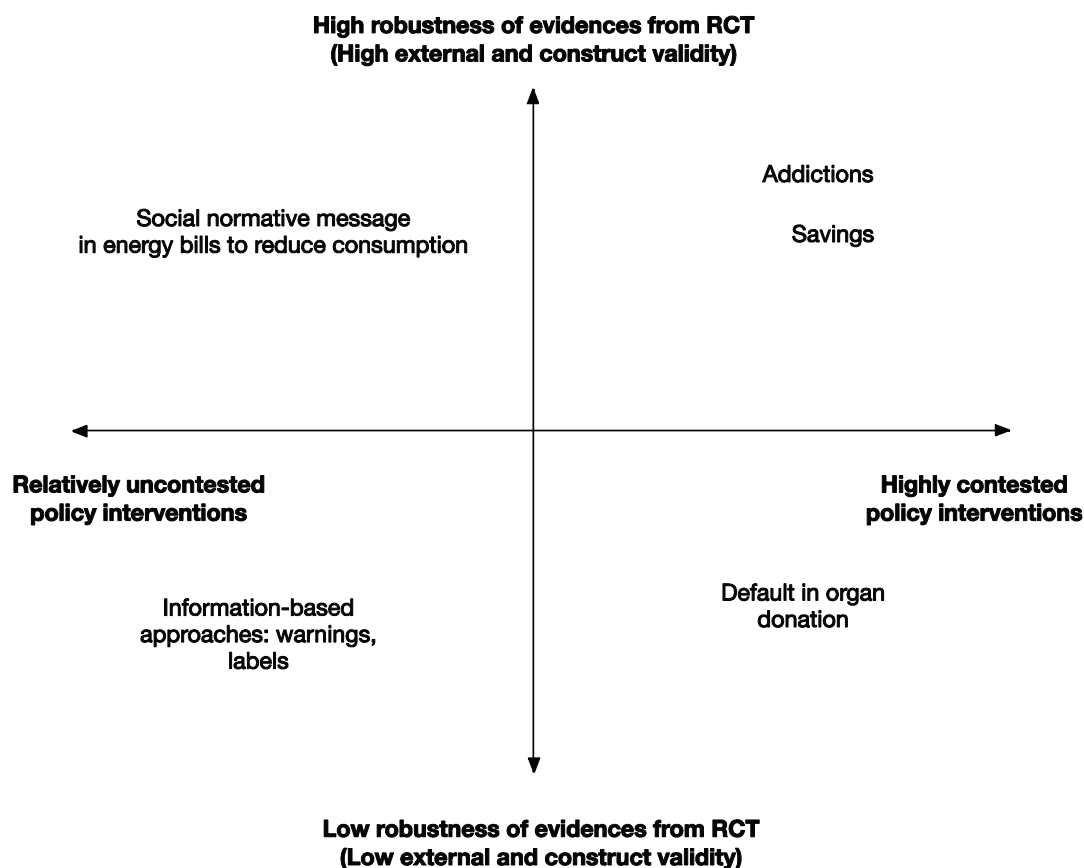
4. Implications for consumer policy

Some lessons can be drawn from the above discussion for consumer protection policy. We can try to map the complexity of the subject into a simple two dimensional taxonomy, mirroring the discussion presented in the two sections above. On the one hand, one could identify a first component related with the high/low external and construct validity of the evidence provided by RCT. We mean all those interventions for which three conditions are met: (a) it is easy to recruit a representative sample; (b) it is easy to design a choice architecture which is ecologically valid, namely understandable by the subjects and mimicking the real phenomenon which is analyzed; (c) the treatments which are tested are valid proxies of the interventions that can be concretely implemented (i.e. are good at

inferring causal parameters); in other words, it is reasonable to assume that a concrete implementation of the measure won't alter dramatically the perceived benefit and cost of the treatment and thus would shift behaviour in line with the experimental evidence.

On the other hand, there is a dimension related with the normative perspective. A certain intervention can be perceived as highly paternalistic, intrusive or simply not being supported by a largely majoritarian consensus in the population. At the opposite a policy option can be perceived as highly technical and thus irrelevant for the majority of the population, not very intrusive or simply being supported by a very large consensus.

The taxonomy is illustrated in Figure 1 below, with some examples.

Figure 1. A taxonomy of behaviourally informed interventions

High robustness of evidences from RCT
(High external and construct validity)

Addictions

Savings

Social normative message
in energy bills to reduce consumption

Relatively uncontested
policy interventions

Highly contested
policy interventions

Default in organ
donation

Information-based
approaches: warnings,
labels

Low robustness of evidences from RCT
(Low external and construct validity)

Source: Authors' elaboration

Página16

In the North West quadrant we identify those sub-domains in which it is relatively easy to meet the three requirements for construct and external validity of RCT. For example, recent evidence show that the default setting of electronic order sets improves doctors prescribing behaviours (Bourdeax, Davies, Thomas, Bewley & Gould, 2013; Harewood, Clancy, Engela, Abdulrahim, Lohan, & O'Reilly, 2011; Jacobs, B. R., Hart, K. W., & Rucker, 2012). In this case, it is easy to get the intervention accepted because it is very cheap, because energy efficiency and environmental protection are usually declared as valuable, as confirmed by survey evidence (Codagnone, Bogliacino, & Veltri, 2013). At the same time the evidence for these interventions has been provided through field experiments, since implementation is pretty straightforward (Costa & Kahn, 2013).

In the North East quadrant we isolate those cases where the evidence from RCT is difficult to contest but the objections against the interventions as highly paternalistic are very likely to be raised. Examples are policies against smokers or directed towards subscription of (medical or other) insurances, or increasing saving rate related with pensions. The latter is actually the first case of successful implementation of nudge, and where the very concept of libertarian paternalism has been conceptualized (Thaler & Sunstein, 2008). In this case, it is possible to design RCTs with almost perfect external validity, e.g. experiments for cigarettes pack purchases or implementation of pension plans at company level, but of course the regulation of health sector, pension and saving decisions, and individual consumption decisions is usually criticized by pro-market thinkers. In the US the Court of Appeals for the District of Columbia has blocked the introduction of pictorial warnings on grounds of violation of the First Amendment's prohibition on government-compelled speech (Bayer, Johns & Colgrove, 2013).

In the South West quadrant we place those interventions weakly contested but for which RCT are difficult to design in an effective way to provide strong evidence. Examples of this domain are those related with information provisions from a behavioural perspective, e.g. envisioned to let shrouded attributes to emerge. These cases are not very debated because in the end information provision is the standard policy implication of Neoclassical Economics. However, in many cases these mechanisms are related to complex purchases (e.g. household

level decisions, very expensive and thus not incentivized in the lab) for which RCT are difficult to design (Costa & Kahn, 2013).

Finally, in the South East quadrant we can identify those subjects for which libertarian paternalism intervention is more likely to face problems to go through the political process to be approved. This is the case for example of default options for organ donors. This may be contested on religious grounds by part of the population and in general is perceived as violating freedom of choice by libertarian and pro-market thinkers (which typically propose a market for organs solutions) (Becker & Elías, 2007). At the same time RCT in this case are difficult to design, in fact the evidence supporting it is coming from observational data and from non incentivized surveys with split ballots for which the risk of Social desirability bias in response is very likely (Johnson & Goldstein, 2003).

5. Final Remarks

Grounding theories of human behaviour on more realistic assumptions is certainly a fundamental issue if one wants to increase effectiveness of interventions. Rigid counterfactuals protocols are instrumental to this process, but are not a magic bullet. Experiments are highly localized evidence, while the intervention itself and the behavioural assumptions aim at being very general. Moreover, experiments deal with means, not ends: they provide insights over how to accomplish a target but cannot determine which aim to prefer.

In this article we give a sketch of the perspectives and challenges that may be raised on BE interventions in various domain, but the main message is that there is no magic bullet.

Nothing in the behavioural science tells us that policymaking will become a simpler matter, or that it is generally optimal to restrain interventions or to use lean regulation instead of structural reforms (Bernheim & Rangel, 2009).

Policy will remain a domain of contrast among interests, but providing more robust evidence will improve transparency, this is way the behavioural turn should be welcome.

References

Barr, M., Mullainathan, S., & Shafir, E. (2013). Behaviorally Informed Regulation. In E. Shafir (Ed.), *The Behavioural Foundations of Public Policy* (pp. 441-461). Princeton, NJ: Princeton University Press.

Bayer, R., Johns, D. & Colgrove, J. (2013). The FDA and Graphic Cigarette-Pack Warnings - Thwarted by the Courts. New England Journal of Medicine 369, 206-208.

Becker, Gary & Elías, J. (2007). Introducing Incentives in the Market for Live and Cadaveric Organ Donations. Journal of Economic Perspectives, 21(3), 3-24.

Bernheim D.B. and Rangel A. (2009). Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. Quarterly Journal of Economics, 124 (1), 51-104.

Behavioural Insights Team. (2011). *Applying behavioural insight to health*. London: Cabinet Office.

Blumenthal-Barby, J. & Burroughs, H. (2012). Seeking Better Health Care Outcomes: The Ethics of Using the "Nudge". The American Journal of Bioethics, 12(2), 1-10.

Bourdeaux, C., Davies, K., Thomas, Bewley, J, & Gould, T. (2013). Using 'nudge' principles for order set design: a before and after evaluation of an electronic prescribing template in critical care. BMJ Quality & Safety.

Camerer, C. F. (1995). Individual decision making. In A. R. J. Kagel (Ed.), *The Handbook of Experimental Economics* (pp.587-704). Princeton: Princeton University Press.

Camerer, C. F. (2003). *Behavioral Game Theory*. Princeton: Princeton University Press.

Camerer, C. & Loewenstein, G. (2003). Behavioral Economics: Past, Present, Future. In C. Camerer, G. Loewenstein & M. Rabin (Ed.), *Advances in Behavioural Economics* (pp.3-52). Princeton: Princeton University Press.

Campbell, D., & Stanley, J. (1963). *Experimental and Quasi-Experimental Evaluations in Social Research*. Chicago: Rand McNally.

Campbell, D. (1969). Reforms as experiments. American Psychologist, 24(4), 409-429.

Codagnone, C., Bogliacino, F., Veltri, G. (2013). Testing CO2 Car labelling options and consumer information. Office for Official Publications of the European Commission.. Luxembourg

Costa, D. & Kahn, M. (2013). Energy Conservation 'Nudges' And Environmentalist Ideology: Evidence From A Randomized Residential Electricity Field Experiment. Journal of the European Economic Association, European Economic Association, 11(3), 680-702.

Codagnone, C. (2007). Measuring eGovernment: Reflections from eGEP Measurement Framework Experience. European Review of Political Technologies, 4, 89-106.

Codagnone C, Veltri GA, Lupianez-Villanueva F, Bogliacino F. (2014). The challenges and opportunities of 'nudging'. Journal of Epidemiology & Community Health, 68, 909–911.

Cronbach, L., Ambron, S., Dornbusch, S., Hess, R. , Hornik, R., Phillips, D., Walker, D., Weiner S. (1980), Toward reform of program evaluation. San Francisco: Jossey-Bass.Dolan,

P., M., H., D., H., King , D., & Vlaev, I. (2010). MINDSPACE: influencing behaviour through public policy. London: Cabinet Office, Institute of Government.

Debreu, G. Excess-demand functions, «Journal of Mathematical Economics», 1: (1974), pp. 15–21.

Fischhoff, B. & Eggers, S. (2013). Questions of competence: the duty to inform and the limits to choice. In: Shafir E, (Ed.). *The behavioural foundations of public policy* (pp. 217-30). Princeton, NJ: Princeton University Press.

Grether, D. & Plott, C. (1979). Economic theory of choice and the preference reversal phenomenon. American Economic Review, 69(4), 623-638.

Heckman, J.J. (2010) Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy Journal of Economic Literature 48: 2, 356–398

Harewood, G. C., Clancy, K., Engela, J., Abdulrahim, M., Lohan, K., & O'Reilly, C. (2011). Randomised clinical trial: a 'nudge' strategy to modify endoscopic sedation practice. Alimentary Pharmacology & Therapeutics, 34(2), 229-234.

Hume, D. (1739). A Treatise of Human Nature. London: John Noon.

Jacobs, B. R., Hart, K. W., & Rucker, D. W. (2012). Reduction in Clinical Variance Using Targeted Design Changes in Computerized Provider Order Entry (CPOE) Order Sets: Impact on Hospitalized Children with Acute Asthma Exacerbation. Applied clinical informatics, 3(1), 52-63.

Johnson & Goldstein (2003). Do Defaults Save Lives?, Science, 302,

Kahneman, D. (2011). *Thinking fast and slow*. London: Penguin Books.

Kahneman, D. (2013). Foreword. In E. Shafir (Ed.), *The Behavioural Foundations of Public Policy* (pp. 7-9). Princeton, NJ: Princeton University Press.

Loewenstein, G. & Prelec, D. (1992) Anomalies in Intertemporal Choice: Evidence and Interpretation. Quarterly Journal of Economics,107(2), 573-597.

Mantel, R. On the characterization of aggregate excess-demand, «Journal of Economic Theory», 7(1974), pp. 348–353. See also the introduction in Bewley, T. (2007) General Equilibrium, Overlapping Generations Models, and Optimal Growth Theory. Harvard University Press.

Martini, A. (2009). How counterfactuals got lost on the way to brussels. In A. Fouquet & L. Méasson (Eds.). *L'évaluation Des Politiques Publiques en Europe, Cultures Et Futurs: Policy and Programme Evaluation in Europe, Cultures and Prospects*. Paris: l'Harmattan.

Micklitz, H., Reisch. L. & Hagen, K. (2011) An Introduction to the Special Issue on "Behavioural Economics, Consumer Policy, and Consumer Law". J Consum Policy, 34, 271–276.

Mullainathan, S. & Shafir, E. (2013) *Scarcity: Why Having Too Little Means So Much*. Times Books

Niskanen, W. (1971). *Bureaucracy and Representative Government*. Cheltenham: Edward Elgar.

Oullier, O., & Sauneron, S. (2010). *Improving public health prevention with behavioural, cognitive, and neuroscience*. Paris: Centre d'analyse stratégique.

Patton, M. (1997). Utilization-focused evaluation: The new century text. Thousand Oaks, CA: Sage Publications.

Pawson, R. & Tilley, N. (1997). *Realistic evaluation*. London: Sage Publications Ltd.

Posner, E. (2001). Controlling agencies with cost-benefit analysis: a positive political theory perspective. University of Chicago Law Review, 68(4).

Saltelli, A., D'Hombres, B., Jesinghaus, J., Manca, A., Mascherini, M., Nardo, M., & Saisana, M. (2011). Indicators for European Union Policies. Business as Usual? Social Indicators Research, 102(2), 197-207. doi: 10.1007/s11205-010-9678-4

Shadish, W., Cook, T. & Leviton, L. (1991). *Foundations of Program Evaluation*. Beverly Hills, CA: Sage.

Shadish, W., Cook, T. & Campbell, D. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company.

Slovic, P. (1995). The construction of preferences. American Psychologist, 50, 364-371.

Sonnenschein, H. (1972). Market excess-demand functions. «Econometrica» 40(3), 549–563.

Stanovich, K. & West, R. (2000). Individual differences in reasoning: implications for the rationality debate? [Comparative Study]. The Behavioral and brain sciences, 23(5), 645-665; discussion 665-726.

Szyszczak, E., (2006). Experimental Governance: The Open Method of Coordination. European Law Journal, Vol. 12, No. 4, July 2006, pp. 486–502.

Thaler, R. (1980). Toward a positive theory of consumer choice. Journal of Economic Behavior and Organization, 1, 36-60.

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

Thaler, R., & Sunstein, C. (2008). *Nudge: improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.

Thompson, V. (2009). Dual-process theories: A metacognitive perspective. In J. Evans & K. Frankish (Eds.), *In two minds: dual processes and beyond* (pp. 171-195). Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. 185, 1124–1131.

van Bavel, R., Herrmann, B., Esposito, G., & Proestakis, A. (2013). *Applying Behavioural Sciences to EU Policy-making*. *JRC Scientific and Policy Reports*. *Luxembourg*: Publications Office of the European Union..

Visser, R. (2003). Trends in program evaluation literature: The emergence of pragmatism: Texas Center for Adult Literacy, Occasional Research Paper No. 5, Retrieved August 29, 2011, from http://www-tcall.tamu.edu/orp/orp5.htm.

Von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. *Princeton*. Princeton: Princeton University Press.

Zizzo, D.(2010) Experimenter demand effects in economic experiments, Experiemntal Economics, 13, 75–98.

Zizzo, D. J. (2013) Claims and confounds in economic experiments, Journal of Economic Behavior & Organization, 93, 186– 195