The Curious Case of Choice Architect: Examining the Philosophical Inconsistencies of Libertarian Paternalism

Kuriakose, Francis and Kylasam Iyer, Deepa

Erasmus University Rotterdam, University of Cambridge

24 April 2017

The Curious Case of Choice Architect: Examining the Philosophical Inconsistencies of Libertarian Paternalism

Francis Kuriakose and Deepa Kylasam Iyer

April 2017

Abstract

Classical economics works on the principle that individuals are rational and make decisions to maximize their self interest. However in real situations, individuals face a conflict between rational and irrational selves leading to decision making that does not leave them better off. Libertarian paternalism proposes a solution to this rationality problem in an individual by conceiving a choice architect. Choice architect is a third party capable of arriving at what a perfectly rational choice would be and 'nudges' an individual towards making that choice. Libertarian paternalists claim that choice architect does not interfere with the freedom of an individual because the choices he offers are easily reversible, i.e, an individual can reject it at any given point in time. Libertarian Paternalism seems to offer the third way between absolute autonomy of individual choice (libertarianism) and third party intervention (paternalism). This paper argues that the conception of a choice architect comes out of a hasty commitment to reconciling libertarianism and paternalism by placing perfect rationality and autonomy in two separate individuals in the case of a single decision making process. The paper proposes alternatives to confront the rationality problem.

JEL: A12, D03, D81, D83

Keywords

Libertarian paternalism, Nudging, Behavioural economics, Choice architect, Rationality, Autonomy

Introduction

Classical libertarians and paternalists look at individual as a decision maker from radically opposing perspectives. Classical libertarians argue that an individual ought to take complete and total responsibility for his actions. Paternalists, on the other hand, assert that it is impossible for an individual to choose what is intrinsically or instrumentally good for him. Dworkin (1972) defines paternalism as follows, 'By paternalism I shall understand roughly the interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests, or values of the person being coerced'. Paternalists, therefore, support third party intervention in individual decision making even against an individual's will under certain conditions. Such justification is based on the assumption that the benefit of interference accrues to the same individual whose freedom has been tampered with. Between the extremes of libertarianism and paternalism, there have been arguments about the possibility of a third way. Libertarian paternalists claim to offer that third way.

Earlier examination of human decision making had analysed the conflict of two 'selves' within individuals at the time of decision making- an intuitive self that takes quick decisions and a rational self that takes calculated decisions (Kahneman, 2002). The most notable advocates of libertarian paternalism, Richard Thaler and Cass Sunstein, have based the justification of libertarian paternalism on the cognitive inconsistencies of individuals in Kahneman's model. Thaler and Sunstein (2008, p.175) define libertarian paternalism as 'the set of interventions aimed at overcoming the unavoidable cognitive biases and decisional inadequacies of an individual by exploiting them in such a way as to influence his decisions (in an easily reversible manner) towards choices that he himself would make if he had at his disposal unlimited time and information, and the analytic abilities of a rational decision-maker.' In other words, the central arguments of the libertarian paternalistic program are that (i) individuals display systematic cognitive deficiencies (ii) these cognitive deficiencies can be manipulated by third party interference that Thaler and Sunstein called 'choice architecture' (iii) choice architecture enables individual to make decisions that they would have themselves taken had they been perfectly rational, and (iv) these decisions once taken are easily reversible, i.e., the individual can reject the decision at any given point in time.

This paper explores the philosophical foundations of libertarian paternalism. What are the core philosophical arguments that drive the major premises of libertarian paternalism, namely

the presence of choice architect and easy reversibility of choices? This paper argues that Thaler and Sunstein model reflects a hasty commitment to libertarian paternalism without probing alternatives. Their model attempts to conciliate perfect rationality and autonomy in two different individuals in the case of a single decision making process by the introduction of a choice architect. The paper points out the philosophical inconsistencies of libertarian paternalism and explores other possibilities to confront the 'rationality problem' based on Kahneman's findings.

Antecedents of libertarian paternalism

Classical libertarian position is reflected in the utilitarian thinking of Jeremy Bentham and John Stuart Mill. Utilitarians have argued the rationale of permitting absolute individual autonomy in decision making. Consequently, they were also absolute on their terms of anti-paternalism. According to Mill, the freedom to choose is a good that was worth in itself compared to the wisdom of others or nature of the choice. He justified his stand on autonomy and individual agency thus, 'A man's mode of laying out his existence is the best, not because it is best in itself, but because it is his own mode. . . .It is the privilege and proper condition of a human being, arrived at the maturity of his faculties, to use and interpret experience in his own way' (Mill, 1869).

Even then, the perspective that paternalism can be allowed in special circumstances is not particularly new within the classical libertarian perspective. Hospers (1980) traces the evolution of libertarianism that accommodates paternalism to various degrees in the development of law. According to him, in the eighteenth century, Jeremy Bentham himself categorised all laws into three broad kinds [i] laws that are designed to protect an individual from harm from others [ii] laws that are designed to protect an individual from harm induced by himself, and [iii] laws that are designed to cause benefits to others. The second category of laws was called 'paternal' laws because the state was attempting to mitigate harm done by an individual against his will and for his own good. This was the beginning of recognising legal paternalism in the utilitarian tradition.

In due course, libertarians accommodated some categories of individuals on whose behalf some third party intervention was desirable in decision making. Children and infants were the first category of individuals who were thought of as incapable of assessing the full impacts of their own actions. The second category of people was the elderly who required help to

examine choices and take decisions made on the basis of the merits of choices. The third category of people was the mentally incompetent who was thought of deserving third party intervention in order to prevent harm to themselves and others. However, the definition of who was mentally incompetent was ambiguous. Yet, what was emerging was consensus that an end-based test, in this case 'harm to one-self and others' that could be an effective criterion to justify third party intervention. Legally, the term 'harm' was variously defined to include 'bodily injury, damage to or theft of property and violation of contract' (Hospers, 1980).

Even then, the question of ordinary normal adult and the possibility of exercising power over their decisions in certain conditions were open. Mill modified his stand on absolute anti-paternalism by stating exceptions in two extreme instances- (i) in cases where an individual caused harm to others because of his decisions and (ii) in the case of a decision that exterminated liberty itself. Mill (1869) decisively wrote *On Liberty* that 'The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right'.

Furthermore, Mill made an exception to paternalism in another case- that of preserving liberty itself. His argument was that liberty had to be preserved at all times. The basic premise here was that if liberty was available at time t_1, then the libertarian stand was that it had to be preserved at all times t_2,t_3, t_4...t_n.. Mill gave two illustrations to make his point. The first was an instance when a person was selling himself off as a slave. He said that when it '..is consideration for his liberty. . . . By selling himself for a slave, he abdicates his liberty; he foregoes any future use of it beyond that single act. He therefore defeats, in his own case, the very purpose which is the justification of allowing him to dispose of himself. ...The principle of freedom cannot require that he should be free not to be free. It is not freedom to be allowed to alienate his freedom.' Another example where third party interference was justified by Mill was when a person was about to take his life. The argument was that this option would put an end to choices in the future.

Therefore, some degree of paternalism within libertarian strand was tolerated toward ordinary individuals under certain conditions which were later referred to as 'harm principles'.

Eventually not just the end-test of justifying paternalism, but the means-based criteria required qualification. Some parameters that were considered in this regard were [i] non coercion or voluntariness of the individual [ii] educated and informed consent of the individual concerned and [iii] healthy psychological state of the individual while making the decision. These criteria read very differently when defined narrowly. For example, the criterion of non-coercion can variously be defined as absence of threat, pressure or any kind of influence depending on how broad the scope of intervention is.

What is interesting in the evolution of libertarian thought was that by paying attention to 'means', the way 'ends' were perceived also changed. It was seen that there was a difference when paternalism was advocated when the third party (say, state) acts on an individual's will to achieve its goals for him and when the party intervened to achieve the goals he would have set for himself. It is the intervention of the latter kind that libertarian paternalism of Thaler and Sunstein aim for. How they came about the particular version of libertarian paternalism requires close scrutiny.

Confronting the rationality problem

Before discussing the unique propositions of libertarian paternalism, it is important to understand what drives the model in the first place. Modern economic thinking rests on the assumption that individuals who make decisions are perfectly rational. A perfectly rational individual has unbounded time, information and will power to make decision that are self-regarding and would maximise his self-interest. A rational individual's preferences among an array of choices are both revealed and consistent. However, in reality, individuals make decisions that are inconsistent with maximising their self-interest because of imperfections in their rationality- absence of unbounded time, information and will power. This 'rationality problem' in an individual has engaged economic thinking extensively in modern times.

The measurement of maximising self interest is done in terms of 'utility'. The term 'utility' has had two distinct meanings in economic thinking. It was Jeremy Bentham who first used 'utility' as the driving force behind individual action. Kahneman (2011) traces the Benthamite utility and concludes that according to Bentham, human actions were derived out of commitment toward two sovereign masters- pain and pleasure- and they dictate what individuals ought to do (ideal choices) and what individuals shall do (real choices). Rational choice theory uses 'utility' in an entirely different sense. According to them, utility is what

rationality dictates an individual to do, i.e., the choice that maximizes his self-interest and that which makes him better-off. It is implied that a rational individual would also want what he enjoys and would enjoy what he chooses, and in this way, also satisfies utility in the Benthamite hedonistic sense. Thus, Kahneman (2011) distinguished Benthamite utility as 'experiential utility' and that defined by rational choice framework as 'decisional utility'.

Kahneman further argued that the cognitive inconsistency within an individual came out of two 'systems' of thinking. Libertarian paternalism is based entirely out of acknowledging this cognitive inconsistency and attempting to correct it with the help of a choice architect. Thaler and Sunstein (2008) take this idea of two systems of thinking to explain decisions that are irrational and justify the introduction of choice architect. In order to do this, they adopt the two systems of thinking proposed by Kahneman as system I and system II types. System I thinking is the automatic system of 'uncontrolled, effortless, associative, fast, unconscious and skilled' decision making while system II thinking is the reflective system of 'controlled, effortful, deductive, slow, self-aware, rule following' decision (Thaler & Sunstein, 2008, p.20).

Thaler and Sunstein then explain the interplay between two systems of thinking as two distinct 'selves' within the same person. Rebonato (2013) brings out the conundrum that libertarian paternalists face at this point. Similar to the externality problem in classical economics, libertarian paternalists face an 'internality' problem. Externality is when the action of an individual causes damage to another individual for which he does not get adequately compensated. Similarly, internality is when system I thinking overpowers system II thinking leading to preferences that are not perfectly rational. The implication of internality is that individuals take decisions that might not be in their self interest. Faced with the intrapersonal comparisons of utility of the two different selves, libertarian paternalists choose to prefer the choices that a rational self (system II thinking) would make. One of the philosophical justifications of 'nudging' through the choice architect is through what the libertarian paternalist describes as overcoming this internality problem.

Unique propositions of libertarian paternalism

The need for a choice architect in libertarian paternalist model comes out of privileging rational self over the less rational self. Furthermore, the external choice architect is forced to make these choices on behalf of system II. This is because a libertarian paternalist is not in a position to ask the individual freely what his preference would be because revealed

preference does not lead to preferred outcomes in the presence of system I thinking. The possibility of observing the individual while making decisions is also ruled out because it interferes with the neutrality of the choice architect. Therefore, the attempt of the choice architect is to ultimately reach the preference of a perfectly rational individual.

Having come to the that choice architect is inevitable and has to force rational choice on an individual decision maker, libertarian paternalist is left with little choice but to work at the libertarian part to make the entire project palatable. Libertarian paternalists achieve this by arguing about the easy reversibility of choices set for an individual. The easy reversibility of decisions presented in choice architecture also means that it is possible to reject the course offered in a choice architecture. Libertarian paternalists claim that easy reversibility makes choice architecture equivalent to not curtailing individual autonomy in decision making.

Thus, there are two propositions implied in the libertarian paternalist model. One is the implicit privileging of the rational self over the less rational self. The reasons for preferring rational self are the attributes perfectly rational individuals have - that of revealed consistent preferences that maximizes self-interest. Second, is the assumption of consistent decisional irrationality in individuals that would make choice architect necessary and effective. Rebonato (2013) brings out these conditions where choice architect would be able to intervene. To begin with, there must be a decisional irrationality that displays regularities whose outcome is necessarily bad for the individual. It must be possible for the choices to be steered towards a different outcome. The choice architect must be able to decipher what the individual would ultimately prefer, inclined to make a 'nudge' that guides the individual towards this preference and that would be better than what a rationally bound agent would obtain. In the absence of these conditions, choice architect cannot perform what he is mandated.

Alternatives from Kahneman's vantage point

The problem of the choice architect, who is forced to conciliate the value of perfect rationality while giving autonomy to individual decision maker, is one of hasty compromise reached by Thaler and Sunstein while proposing libertarian paternalism. For a moment, accepting the two systems of thinking proposed by Kahneman and a prima facie commitment that rational choices are better than other kinds of choices, what are the possible methods to overcome the problem of cognitive inconsistency that lead to less-than-rational choices of individuals?

One way to think about the problem is to return to the vantage point of Kahneman and closely examine the various types of cognitive deficiencies that he illustrates. There are two directions available here. One is to think through minimizing the cognitive deficiencies through information networks. Dawes (1971) had proposed that introduction of linear model of choices rather than an array of choices help individuals in making choices that are consistent with the preferences of system II thinking. In another instance, partial 'debiasing' of errors has been achieved by group decisions than individual decision making in various situations (Tversky & Kahneman, 1974). The process of viewing options jointly than separately and understanding the underlying principles than surface assumptions greatly encourage moving toward a system II type of thinking. Fischhoff (1982) reviewed four strategies to overcome decision making bias- offer warning about the possibility of bias, offer direction of the bias, offer feedback loops to correct bias, and extended training to minimize the decisional bias. Another well known example of cognitive bias is through 'framing effect' in which individuals respond differently to a choice based on the way they are framed. Experiments have proven that framing effects can be greatly diminished while observed in social context and through providing external credible advice (Druckman, 2001). Therefore, a number of information network strategies might be effective in reducing decision making bias without compromising on the transparency of the process.

The second set of alternatives available is to reconfigure the understanding of rationality itself. Sen (2004) describes rationality as the need to subject one's choices to the demands of reason. This definition significantly moves from the understanding of rationality as self regarding, utility maximizing choices that expected utility theory defines. Sen's definition of rationality involves individual's scrutiny of goals (values and wants) that might be acting out of self-interest, commitment (doing something right) or altruistic ends (doing something for others). Making the definition of rationality broader can reconcile system I and system II types of thinking and also the conflict between values of autonomy of libertarians with intervention of paternalists. According to a perspective that identifies rationality as the test of reasonableness, rationality *is* freedom.

Conclusion

Cognitive deficiencies of individuals have been identified as one of the reasons for deviation of preferences from the predicted responses of expected rational utility theory. Thaler and Sunstein proposed the presence of a choice architect who would nudge an individual toward

choices that he himself would have taken had he been perfectly rational. The presence of an external choice architect intervening with the autonomy of individuals is overcome by the proposition that the choices presented are 'easily reversible'. But a choice architect can provide easily reversible choices only under certain conditions that are difficult to meet. This paper argues that there are alternatives available to rectify the cognitive inconsistencies of an individual decision maker rather than resort to the uneasy compromise that libertarian paternalists propose.

One set of alternatives looks at providing information network to an individual decision maker to make better choices. Presence of information, linear modelling of presentation, feedback loops, training, credible advice and group decision making diminishes the effects of cognitive bias to a large extent. The second set of alternatives looks at rationality as a test of reasonableness embracing the ideas of conflicting values (self regarding/other regarding) in individual decision making.

The cost of cognitive deficiency in decision is increasing due to the impact of decisions on private and public resources. While thinking of correcting systematic deficiencies, it is important to understand that new designs would be implemented in political economies where accountability and transparency of the system is as valuable as the end results of decision making. Confronting the rationality problem requires renewed thinking in the emerging contexts.

References

Dawes, R.M. (1971). A case study of graduate admissions: Application of three principles of human decision making. American Psychologist, 26 (2), 180-188.

Druckman, J.N. (2001). Using credible advice to overcome framing effects. *The Journal of Law, Economics and Organization*, 17 (1), 62-82.

Dworkin, G. (1972). Paternalism. *The Monist*, 56 (1), 64-84.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Hospers, J. (1980). Libertarianism and legal paternalism. *The Journal of Libertarian Studies*, 4(3), 255-265.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahneman (Eds). *Heuristics of Intuitive Judgment: Extensions and Applications*. New York: Cambridge University Press.

—— (2011). *Thinking, fast and slow*. New Delhi: Penguin Allen Lane.

Mill, J.S. (1869). *On Liberty*. London: John W. Parker & Sons.

Rebonato, R. (2013). A critical assessment of libertarian paternalism. Retrieved 20 September 2016 from http://s3.amazonaws.com/academia.edu.documents/37911818/version2.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1486024306&Signature=DbwAvkWfx75cORT6xGAloWYVI50%3D&response-content-disposition=inline%3B%20filename%3DA_Critical_Assessment_of_Libertarian_Pat.pdf

Sen, A. (2004). *Rationality and freedom*. Cambridge: Harvard University Press.

Thaler, R.H. & Sunstein, C.R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. New Delhi: Penguin Books India.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.