An Anselmian Case Against Libertarian Paternalism

Xavier Meulders
Hasselt University

Abstract. Libertarian paternalism is "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 her decisions (in an easily reversible manner) towards choices that she herself would make if she had at her disposal unlimited time and information and the analytic abilities of a decision-maker." Hence, the rationale behind libertarian paternalism is pragmatic rather than purely academic. Libertarian paternalism seemingly operates under the banner of freedom. However, it fails to make its (metaphysical) presuppositions explicit, some of which are problematic. Particular attention should be paid to libertarian paternalism's endorsement of a "two selves" picture of human rationality. This picture is fundamentally mistaken and leads to a misconception of freedom. A non-dualist account of freedom that has been formulated by Saint Anselm of Canterbury (1033–1109) could offer a way out of this conundrum.

Keywords: libertarian paternalism, economic rationality, behavioral economics, free will, Susan Wolf, Anselm of Canterbury

> "*Advantage! What is advantage? And will you take it upon yourself to define with perfect exactitude precisely what man's advantage consists in? And what if it so happens that on occasion man's profit not only may but precisely must consist in sometimes wishing what is bad for himself and not what is advantageous?*" Fyodor Dostoevsky, *Notes from the Underground*

1. Libertarian paternalism: key concepts and a general critique

Classical paternalism used to emphasize an individual's inability to choose what is either *instrumentally* good as a means to some given end or what is *intrinsically* good by virtue of some overarching idea of the good life and/or society's common interest. Nowadays, however, this paternalist premise is at odds with contemporary beliefs in individualism, personal autonomy, and value pluralism that are part and parcel of most, if not all, Western liberal democracies. With regard to value pluralism, it may nowadays even be safe to assert that any substantive talk about intrinsic value has become altogether obsolete.

As a result, paternalism has had to restyle itself in a less condescending version, leaving behind the often pedantic rhetoric surrounding the alleged existence of the intrinsically valuable. The phoenix that arose out of the old paternalism also has a name: *libertarian* paternalism. It is the adjective that makes all the difference of course, since libertarian paternalism claims that it is able to reconcile the paternalist's concern for the improvement of an agent's choices from a first person's perspective *thereby leaving his individual freedom and autonomy intact*. It is the latter addendum that distinguishes libertarian paternalism from its classical counterpart. But what exactly is libertarian paternalism? As the economist Riccardo Rebonato in his excellent critical overview explains, libertarian paternalism can best be understood 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 her decisions in an easily reversible manner towards choices that she herself would make if she had at her disposal unlimited time and information" (Rebonato 2012, 6).

This needs some conceptual clarification. For a start, libertarian paternalism distinguishes itself from classical paternalism in that it is *only* concerned with what is *instrumentally* good, that is, with the *means* deployed by a human agent. In contrast with classical paternalism, its libertarian cognate remains silent about the *ends* that one has to pursue in life.

By contrast, libertarian paternalism rejects the standard view of economic agency as fully rational. Libertarian paternalism asserts that human beings are prone to decisional irrationalities. Their choices are frequently affected by apparently irrelevant framing. Those choices are inferior since they lead to suboptimal outcomes that the agent would not have chosen if they were perfectly rational. It is up to the libertarian paternalist to intervene in an individual's choices through discrete intervention in the choice architecture, i.e., the way choices are presented to them. These interventions are said to be *easily reversible:* the person who experiences themself to be targeted by the intervention must be able to revert or sidestep it at little or no cost.

A simple example might illustrate the attractive lure of libertarian paternalism.[1] As of 2010, almost 100 per cent of the Austrian population had consented to be organ donors. This number drops to a meagre 12 per cent in neighboring Germany. Differences in culture, morals, or political institutions could hardly account for this glaring difference. The difference could only be explained by virtue of a "manipulation of choice architecture" that has been implemented in Austria, but not in Germany. German citizens have to give their explicit consent to become organ donors, whereas their Austrian counterparts have to undertake definite steps to unsubscribe from the scheme. Both German and Austrian citizens thus seem to retain their nominal freedom: neither of them is forced to become an organ donor at a later stage in life. According to the libertarian paternalist's standards, it is the Austrian citizen

[1] See Rebonato (2012, 8-9) for a fuller discussion. The numbers I quote may be outdated by now since the survey was conducted in 2010.

whose choice should be considered to be *truly* free as it reflects his *genuine* interest. Hence, the libertarian paternalist may claim that this discrete intervention properly works: preferences are being rearranged to reflect better outcomes without substantially intervening with personal freedom. After all, even according the most basic and rudimentary accounts, the *easy-reversibility* criterion stands out as a distinguishing, essential feature of libertarian paternalism.

But a closer examination may actually demonstrate that the libertarian paternalist is crying victory way too soon. One may rightly ponder whether the preferences revealed by German and Austrian citizens respectively reveal either the preferences discretely imposed by the choice architect or the preferences revealed by exercising their real freedom as human agents. If the numbers cannot provide us with anything germane about the latter, the easily-reversibility criterion seems to be ill-fated and the whole (methodological) edifice of libertarian paternalism starts to collapse.

It is my contention that the numbers do not reveal anything informative at all. The easy-reversibility criterion states that the default choice imposed by the choice architect must be able to be overturned at little or no cost. However, "cost" is a highly *subjective* notion and tied up with an individual's preferences, values, and choices. Choices are made against the backdrop of several possible alternatives of which only one can be realized. Moreover, choices are also forged within the broader context of personal life plans and value judgments which are always constrained by limited time, information and resources. It is within this broader horizon of both personal and factual circumstances that expected utilities and (expected) costs for foregone alternatives are assigned their respective subjective values. For example, when I have three euros at my disposal, I am able to buy either a loaf of bread or a newspaper. Whatever my choice may be, this actual choice forecloses the other alternative, given limited time and resources. What is more, when I decide to buy the loaf of bread, for example, this choice reflects my actual preference or value judgment with regard to the loaf of bread. For instance, it may satisfy my desire to have breakfast. At the same time, the alternative of buying a newspaper is thereby foregone.

There is nothing particularly mysterious about this example. From an economic point of view, both the actual choice as well as the foregone alternative take place within standard free market conditions. There are no forced or coerced transactions, nor is any transaction bound by a condition. Foregoing buying the newspaper will not result in a financial or other penalty, for instance. It is really "up to the economic agent" to evaluate all the possible outcomes. But when we return to the libertarian paternalist's example of organ donation, it turns out that the choice architect has rearranged the burdens and conditions in such a way that choosing to escape the libertarian paternalist's default option may be a very costly affair. It may be the case, quite plausibly so, that Austrian citizens who are automatically enrolled in an organ donation scheme face steep bureaucratic red tape when withdrawing from the scheme. Or it may be that citizens are inadequately informed about the procedure for

unsubscribing.[2] Whether or not that is the case cannot be inferred from the numbers that the libertarian paternalists cite. These numbers remain silent in the absence of a more stringent interpretation.

This problem poses a methodological dilemma for the libertarian paternalist . After all, there must be a way to gauge an agent's true preferences that are in the best interest *of themselves* so that a neat demarcation could be drawn between subjective preferences *as they actually are* versus those preferences *as they ought to be*. I will elaborate upon this distinction, which anticipates libertarian paternalism's endorsement of a dual self, in much more detail later on. Libertarian paternalism has two possible paths to escape from this dilemma, neither of which seems very promising. The first one is to state that revealed and "true" preferences are actually one and the same thing. But then the whole rationale behind libertarian paternalism collapses: after all, why still worry about alleged cognitive biases and irrationalities if revealed preference theory or even neoclassical economics as a whole is all there is to teach us about human rationality? And wasn't it precisely the aim of libertarian paternalism to dig a way out of the naive anthropology endorsed by neoclassical economics? A second path out of the dilemma might be even more perplexing. It simply consists in a prior design of some general characteristics of what an "ultimate preference" or a "choice in one's best interest" should be. It is then the libertarian paternalist rather than the individual agent himself who determines the goals that need to be achieved. This would be at odds with libertarian paternalism's tacit endorsement of value pluralism. Indeed, libertarian paternalism then simply turns into ends-paternalism. As I will try to show, the libertarian paternalist usually undertakes the first route, only to filter in on the second route after a short deviation.

With regard to the first route, something must be said about revealed preference theory. The original formulation of revealed preference doctrine, developed in the 1930s by Paul Samuelson, assumed a fixed and constant preference scale that is unvarying over time. Revealed preference theory also assumes a couple of auxiliary axioms that are considered to be tantamount to economic rationality, namely the *transitivity* and *completeness* of choices. A person's preferences are transitive when, given a set of choices *x*, *y*, and *z*, this person prefers *x* to *y*, *y* to *z* and therefore *x* to *z*. His preferences are complete if and only if given a set of choices *x* and *y*, that person prefers *x* to *y*, *y* to *x* or is indifferent with respect to the choice between *x* and *y*.[3]

[2] It is for this reason that libertarian paternalism is at times reproached for having an accountability problem. This implies that the means it uses are not fully transparent and cannot be assessed on an objective basis. After all, the libertarian paternalist has a licence to impose any measure whatsoever as long as they can claim that the people targeted by the measure retain their nominal freedom. However, it is one thing to offer employers a tax credit in order to enable their staff to get vaccinated against COVID-19, but is quite another thing to have the whole of society strapped in a near-to-ineluctable system of QR-codes and vaccine passports just so that social life can go on. Isn't libertarian paternalism using other persons as a mere means of achieving its own agenda? This is a more normative criticism of libertarian paternalism that I will leave aside in the development of my own argumentation. For further discussion, see for instance Seymour Fahmy (2018, 96-107).

[3] For further discussion see Hausman and McPherson (2006, 45–59).

Through the application of those auxiliary axioms, revealed preference theory tries to navigate a daring route between the Scylla of mere tautology and the Charybdis of a (philosophically) too substantial definition of utility and preference. The formal prerequisites of rational ("revealed") action say something about the relations between '*x*', '*y*,' and/or '*z*' without needing to fill in the blanks covered by these algebraic arguments. However, this maneuver is bound to fail. Firstly, one may question whether the axioms of transitivity and completeness contribute something essential to the concept of economic rationality. For instance, revealed preference theory assumes that choices remain constant over time. This is a false assumption, because people may effectively adopt new patterns of choice in the light of new goals they want to achieve

Secondly, these auxiliary conditions are assumed to generate the transition from a merely tautological concept of utility towards a more substantial one in terms of rational choices and/or informed preferences. Only alleged rational choices can be choices in the real sense of the word. This is precisely where libertarian paternalism finds a lead. After all, the marriage between libertarian paternalism and revealed preference theory is an uneasy one. At first glance revealed preference theory seems inapt for discriminating between the preferences *as they are* versus preferences *as they should be* under idealized conditions. But since the alternative to revealed preferences would be something akin to *informed* preferences, thereby enticing libertarian paternalism towards the dead-end street of ends-paternalism, this suggestion does not seem very appealing either. Rather reluctantly, libertarian paternalism cannot do otherwise but embrace revealed preference theory—albeit it with a twist. Rebonato explains how:

> Given the supposed pervasiveness and inevitability of these cognitive limitations, they [i.e., the libertarian paternalists] then refuse to make use of revealed preferences. They are therefore faced with an enormous problem of interpersonal intelligibility of preferences. And this is where they have to make a conceptual leap, and implicitly embrace a very restrictive version of utility. As their distrust of revealed preferences makes the mental states of individuals virtually inaccessible, and as informed-preference satisfaction cannot be left to the decision-maker, they attempt to estimate the utility attaching to the course of action that connects the externally visible start and end points of a decision-making process. What do I mean by 'externally visible'? Consider, for instance, a smoker. Presumably, she would prefer to live a longer life. Why is she smoking then? According to the libertarian paternalists, it is extremely unlikely that she has made a rational trade-off between the pleasure of many cigarettes (the visible starting point) and the significantly increased likelihood of suffering from many lethal diseases (the visible end point). She can't, in their eyes, *really* prefer inhaling tobacco to living a longer and healthier life. Since there is supposed to be nothing else in the utility calculus, she must be making a 'mistake.' Perhaps she is poorly informed about facts ... Let's give her information and, if this is not enough,

> let's employ all the context-manipulation and related tricks in the cognitive book to make her make the 'right' decision. By doing so, we must have made her happier, the libertarian paternalists would say (Rebonato 2012, 196).

So instead of having *one* preference scale on their dashboard, the libertarian paternalist actually observes *two*: one belonging to a rational decision maker and another one to an inferior myopic agent. It now becomes obvious why it is that the libertarian paternalist who endorses a modified version of revealed preference theory ultimately has to tacitly adopt ends-paternalism: they are actually concerned with those outcomes of the rational decision maker that *ought* to be fostered. Obviously this is a substantive value judgment that goes without any warrant. It actually betrays circular reasoning: we ought to listen to what individuals would chose if they had ample time to reflect, because this is the decision that would be reached by a person who had been given ample time to reflect.

This circular reasoning stems from libertarian paternalism's reluctance to lay its methodological cards openly on the table. Is libertarian paternalism a descriptive or prescriptive enterprise? From our opening remarks that libertarian paternalism only concerns itself with that which is instrumentally valuable, one might assume that its scientific outcomes are, by and large, descriptive: certain cognitive biases are introduced as facts whereas it is up to policy makers to either adopt or decline its remedies. Hence, libertarian paternalism is often interpreted along value-free lines.

It is my contention that this conception of libertarian paternalism as a value-free science is mistaken because the idea of value-free science *in general* rests on a mistaken premise. Even though this is not the right place for a deep dive into the so-called dichotomy between fact and value, I think it may be worthwhile dwelling on the idea of a value-free science.[4] Since authors working in the tradition of libertarian paternalism remain silent on this broader theme, it is unfortunately impossible to directly engage with the relevant literature at hand. I will therefore have to make recourse to other sources that may, or may not, have had a (methodological) influence on libertarian paternalism's and/or neoclassical economics's core tenets.

The idea of a value-free social science dates back to the nineteenth century when envy of the impressive progress made in the natural sciences led to the idea that all sciences should be guided by the same methodological principles. This idea culminated in the work of the German sociologist Max Weber, who contended that, since facts and values are absolutely heterogenous, social science must be an ethically neutral enterprise. Weber famously introduced the concept of the ideal type, which was supposed to be a value-free construct according to which certain social phenomena could be interpreted as concrete moments of a more general concept (i.e., "type"). Weber defined the ideal type as follows: "An ideal type is formed by the one-sided accentuation of one or more points of view and by the synthesis of

[4] For a critical examination of this distinction, see Putnam (2002, 28–45).

a great many diffuse, discrete, more or less present and occasionally absent concrete individual phenomena, which are arranged according to those one-sidedly emphasized viewpoints into a unified analytical construct" (Weber 1949, 89). But how is the social scientist supposed to accentuate one or more points of view? What points of view are relevant to the social scientist? What license does he have to accentuate one point of view and leave another out of consideration? After all, we might safely assume that the construction of the ideal type does not happen haphazardly, especially given Weber's own criticism of positivism.

Weber's answer to this methodological conundrum is that it is done "by reference to value." Weber is adamant in admonishing his audience that the constitution of social objects by reference to value does not entail making a value judgment on the part of the social scientist. On the contrary, the scientist merely traces those phenomena to their causes, refraining from any substantive value judgment whatsoever. In so doing, Weber skillfully brought the realm of values back to the fore, albeit in a value-free guise: "value" in the Weberian sense turned out to be shorthand for "cause."

Yet, this still begs the question: how is the social scientist supposed to assign value to some aspect over another aspect without making a judgment about which aspect should be assigned a value? Weber claimed an absolute quietism in this regard, refusing to subscribe to any particular value whatsoever. As a result, his apprentice is left empty-handed and without any guidance as to how to strive for knowledge. The question of selecting the relevant "points of view" thus still remains open. At least as far as I can see it, it is question-begging how this operation could be carried out without recourse to making value judgments, that is, to endorse "certain values over references to value."

Perhaps an example may illustrate this. Suppose that a social scientist, for instance a legal scholar, wants to describe a legal system bound by the rule of law. It is probably uncontested that such a legal system entails, among others, the following features: that the law is sufficiently general, that it is publicly promulgated, that it cannot be applied retroactively and that it is possible to obey.[5] These are undoubtedly necessary characteristics of a legal system that abides by the rule of law. However, a necessary condition does not always imply a sufficient one. For instance, in the wake of the Nuremberg trials that were held between November 1945 and October 1946, jurists were shocked to find out that the legal system of Nazi Germany actually *did* conform to all of the requirements of the rule of law. The antisemitic Nuremberg laws enacted in 1935, depriving Jews of their citizens' rights, undoubtedly were a blow in the face of humanity yet they were sufficiently general, publicly promulgated, prospective and possible to obey. It is precisely for this reason that some postwar (German) legal scholars, such as Gustav Radbruch, formulated additional material requirements to the thin conception of the rule of law, such as respect for human rights and minorities, amounting to a more encompassing notion of the rule of law.

[5] See Fuller (1969, 33–94) for a thorough overview.

So which legal system best matches the requirements of the rule of law? Our unfortunate legal scholar, who just acquainted himself with Weber's methodological works, is placed in a dilemma: Will he just describe the "thin" or "formal" concept of the rule of law, or will he prefer a description along broader and more substantive lines?[6] It definitely requires a sound mind that is able to grasp the different values at stake in order to decide which concept should be useful as an ideal type. And even if our legal scholar were to refrain from making a value judgment himself, he still cannot operate in a complete moral vacuum. At the very least, he has to take into account those moral values that arise in a certain social environment—an environment that changes over time and that subscribes to moral values that might evolve in accordance with new historical, religious, or socioeconomic developments. In any case, the decision to attach more importance— "by reference to value" —to one phenomenon over another cannot be an act of neutral description, as this would amount to an infinite regress in which a new description is made about which phenomenon to describe. A description, value-free or not, can only be made after the scientist has decided upon his subject matter, and this requires the execution of at least one value judgment. It is one thing to go unhindered by the external ethical requirements imposed by society upon scientific practice (for instance, to be unmoved whether or not one's findings in theoretical physics will be used to develop nuclear weapons), but as a scientist, one cannot remain blindfold with regard to science's internal ethical requirements that demand unremitting adherence to intellectual integrity and the full elucidation of one's research method, scientific and metaphysical presuppositions, and conclusions by all means available.[7]

Similarly, libertarian paternalism cannot refrain from making value judgments either, and it is for this reason that it cannot be a purely descriptive enterprise. For instance, in accentuating the notion of a preference "in one's best interest," libertarian paternalism must already have committed itself to a value judgment with respect to those ideas. Another example is libertarian paternalism's endorsement of a dual account of decision making.

Indeed, libertarian paternalism assumes there to be two different decision makers. This needs some further unpacking. According to the libertarian paternalist program, one mode of decision making is automatic, whereas the other is reflective. They have been baptized System I and System II respectively (Thaler and Sunstein 2008, 21). System I is considered to be *uncontrolled, effortless, associative, unconscious,* and *present-oriented* whereas System II is thought to be *controlled, effortful, deductive, self-aware,* and *future-oriented*. It is through the interplay of both systems that human agents make their decisions. However, it seems that in real-life situations System I usually takes the lead, thus advancing decisional irregularities. Libertarian paternalism then further assumes that decisional

[6] It is for this reason that continental "civil law" jurisprudence knows at least two terms that describe the two different phenomena whereas the English language only knows the term "rule of law." For instance, in French the notion of *état légal*, more or less corresponding to the thin concept of the rule of law, is often contrasted with that of the *état de droit*, which is a state itself bound by the limits of law. These terms could best be translated as "legal state" and "lawful state" respectively.
[7] For further discussion see van Dun (1986, 17–32).

irrationality displays *regularities*, that these irrationalities are *bad* and that it must be possible for an external choice architect to assess *what choices the agent would actually prefer given an unlimited amount of time and information*. If these requirements are met, the libertarian paternalist could refashion the choice architecture in such a way that better decisions ensue.

As has already been anticipated, libertarian paternalism cannot actually do without neoclassical rationality. It aims at fostering the rationality of the rational decision maker or *homo economicus*. Thaler and Sunstein gave the homo economicus a new name: the Econ, someone who "can think like Albert Einstein, store as much memory as IBM's Big Blue, and exercise the willpower of Mahatma Gandhi."

According to the libertarian paternalist, System I and System II seldom act in harmony with one another. On the contrary: the often irrational caprices ensuing from the shortsightedness of System I stand at permanent odds with the cool reasoning of System II. And any deviation from the golden standard of economic rationality is seen as an error or cognitive failure. This constitutes not merely a descriptive state of affairs but a normative one as well, since it is assumed that System I actually inflicts *harm* on System II. To buttress this claim, libertarian paternalism invented the notion of an "internality" – analogous to the concept of an externality. An externality usually involves third parties that cannot be fully compensated for the negative repercussions of economic activity. The third party is thus an innocent bystander. Similarly, the self-embodied by System II is an innocent victim of the whims of System I.

This juxtaposition of two selves ultimately leads to an inadequate account of reason and logic on the one hand, and an inappropriate conception of the objects we value and choose on the other.

Since libertarian paternalism is interlaced with ideas and concepts borrowed from behavioral economics and cognitive psychology, it might be instructive to turn to some research findings from the psychologists Daniel Kahneman and Amos Tversky. In the 1980s they conducted a series of experiments demonstrating the alleged cognitive deficiencies that occur in accomplishing relatively simple cognitive tasks. One of the experiment's results came to be known as "*the Linda problem*."

Imagine you are told the following about a person named Linda: Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Now, which of the following statements is more probable?

1) Linda is a bank teller.
2) Linda is a bank teller and is active in the feminist movement.

The experiments revealed that a vast majority of the participants assumed the second statement to be more probable. From a statistical viewpoint, however, this does not make much sense. After all, there are more bank tellers *generally speaking* than there are *feminist*

bank tellers. It seems thus that the respondents choosing statement (ii) were lured by the cognitive deficiencies operational in System I. Kahneman concluded that by adding irrelevant context to relevant facts the participants fell prey to the conjunction fallacy. According to Kahneman, the conjunction fallacy occurs when someone judges the simultaneous combination of two possible states of affairs to be more likely than a single one in isolation.

But did Kahneman and the libertarian paternalists along with him really prove that human beings are thoroughly "irrational"? With regard to the Linda experiment, it is not certain whether the participants were behaving irrationally. After all, it may well be the case that the biographical information of Linda was effectively interpreted as relevant at disclosing certain further ambitions of Linda. This, in its turn, may have given way to an entirely different meaning of "probability," one that would deviate from the standard statistical meaning towards something more akin to narrative plausibility. This example apparently reveals that there and here is more to reason than mere logical deduction. Formal logic typically ignores the context of argumentation. It neglects the fact that claims are made and challenged by speakers and listeners. They take them instead as raw data in need of formalization. And yet, at the very least, classical Aristotelian logic cannot do without agreement on universals and genera, for instance, on the fact that "all men are mortal." But these agreements take place within a dialogical horizon that even formal logic cannot entirely avoid.[8]

2. Libertarian paternalism and the real self-view

The Linda experiment may reveal something philosophically deeper about the complicated relationship between System I and System II. Indeed, libertarian paternalism apparently endorses a *two selves picture* of human agency, one that closely ties with the *Real Self View* discussed by moral philosopher Susan Wolf (2005, 258–74). According to Wolf, adherents of the Real Self View usually distinguish between the values, beliefs, and desires of an inferior, alienated self vis-à-vis the values endorsed by a core or "deep" self. However, it is far from obvious which set of values and convictions accrues to the real, authentic self and which one does not. Some, and perhaps the most important things we value in life are handed down to us by our parents, teachers, and friends. To which "self" do those values belong then? To an empirical, unreflective self or to a transcendental pure ego? This is far from clear. Advocates of the Real Self View face two options to circumvent both horns of the dilemma, neither of which is very promising. They might either try to purify the contents of the ideal self in such a way that nothing remains in it, turning the real self into an empty vessel, or they might argue that in order for the real self to emerge, another self is needed to furnish the real self with value-content. But if an even deeper self than the real self is needed to account for it, the spell of an infinite regress looms large, one in which ever-deeper selves need to sustain the

[8] For an excellent treatment (in Dutch) of the dynamic interplay between logic, dialectics and rhetoric see Rutten (2018, 7–62).

upper layers of "real selves." Hence, the Real Self View collapses under the weight of its own inner contradictions.

If Susan Wolf's argumentation is correct—and I think it is—then it might strike a deathblow to the dualist vision endorsed by libertarian paternalism. After all, why should we accept that the self of System II corresponds to a "realer" or "deeper" self than System I? For example, libertarian paternalists commonly claim that individuals suffer from present bias. It is assumed that people who commit themselves in the future to, say, a healthier lifestyle actually refrain from putting their intentions into practice once the future becomes the present. Hence, good intentions risk being postponed indefinitely, at which point the libertarian paternalist will have to step in and take over. But is it necessarily irrational or even bad to have one's current preferences satisfied over future ones? We do not need to endorse David Hume's ironic quip that "it is not contrary to reason to prefer the destruction of the whole world to the scratching of my finger" for us to see that *hindsight bias* may be an equally strong impediment on our choosing and flourishing as human beings. Excessive orientation on the future may lead to underconsumption of desirable activities in the present and even vices such as social apathy. Internalities go both ways.

3. *Omnis volens ipsum suum velle vult*: Anselmian understanding of unified will

So why should we assume that there is anything dignified or morally superior about our abilities as a utility-maximizing, future-oriented welfare unit if we could equally strike a blow for our lesser rational self? Perhaps it is the two selves view that stands in the way of a proper assessment of which self to favor. I think that our discussion might be greatly improved if we involve Saint Anselm's teachings on freedom, will, and choice.

Contemporary discussions of freedom and free will usually invoke the concepts of "ultimate authorship" and "alternative possibilities" to demarcate the argumentative lines of discussion. With regard to ultimate authorship, so-called libertarians will commonly emphasize that it is up to the agent to exercise his will, unfettered by the social, biological, or cultural factors that may otherwise obfuscate his act of free willing. As we discussed in the previous paragraph, advocates of the Real Self View are adamant in their emphasis of ultimate authorship—however, they have not proved themselves very successful so far. As with regard to alternative possibilities, these, too, seem to be required in order to make a cogent argument about freedom: if I weren't able to choose otherwise than I did, my action cannot be considered to be free and I would no longer be liable to be held responsible for my action.

But perhaps the dyad of ultimate authorship and alternative possibilities is a false conceptual framework to begin with. In judging that the area of the square whose side is the hypothenuse equals the sum of the areas of the squares on the other two sides I demonstrate that I have adequately grasped the correctness of the Pythagorean theorem. Even though it is up to my own intellectual capabilities to freely grasp the different geometrical concepts at play, it would be rather awkward to claim that the truth-value of the proposition would be

saturated because of me judging something to be the case. Claiming "ultimate authorship" for cognitive acts would pave the way for the most devastating forms of psychologism. Similarly, and with regard to alternative possibilities, one may think of the Good Samaritan who magnanimously helped a robbed man stranded on the road. Even though we cannot deny that the Good Samaritan was free in doing what he did, wouldn't it be bizarre to say that the Good Samaritan's choice was one out of several "alternative possibilities"? Isn't there something more compelling that arises out of the struggle between right and wrong than randomly selecting one possible world or another?

If this is the case, ultimate authorship and/or alternative possibilities may supply (necessary) conditions for freedom, but actually say nothing at all about the essence of freedom. For Anselm, freedom has a direct normative significance. In book III of *De libertate arbitrii,* Anselm defines freedom as "the ability to keep uprightness-of-will for the sake of this uprightness itself." The Latin word that Anselm uses to denote uprightness is *rectitudo*, which also appears in works like *De veritate* where Anselm uses it to denote the *recta significatio* that correctly connects words to named objects. Rectitude is in the will but not part of the act of willing. To put it otherwise, rectitude denotes the "direction of fit" of our intentional acts. They can be either correct or incorrect, that is, either contain or lack rectitude. Rectitude is thus a quality, standard, or excellence of the will.

Two other important Latin terms that appear throughout Anselm's works are *libertas arbitrii* and *pervelle*. With regard to *libertas arbitrii*, it must be stressed that it denotes the will's capacity to follow the path of rectitude. It should thus be distinguished from *liberum arbitrium*, the act of free willing itself, or to choose amongst several possibilities, and which only becomes actual through the pursuit of uprightness of will. Finally, *pervelle* could best be translated as the perseverance of the will to continuously will rightness of will or *rectitudo* for its own sake. A will that continuously wills rectitude for its own sake is a perfectly free will.

I admit that this terminological interlude may obscure things rather than shed light on them. Let me therefore illustrate Anselm's theory with the aid of an example.[9] Consider two people each of whom has an obsessive habit: a drug addict and Socrates. Both men share certain compulsive character traits. The drug addict wants his cravings to be continuously gratified by having yet another heroin shot, whereas Socrates's monomaniacal behavior reveals itself in his constant and obsessive need to question and understand things. In both cases, their behavior may lead to their own self-destruction and in the case of Socrates, it actually did. From a modern post-metaphysical angle, it would seem that both men are actually under the sway of certain impulses and cravings. The best thing modern philosophy can come up with, then, is a compatibilist account of freedom: urges and cravings that help foster self-construction rather than self-destruction add to freedom. In the worst case, we

[9] I borrow this example from Nash-Marshall (2008).

might be lured to think that both men are equally unfree because they lack ultimate authorship.

But this misses the gist of Anselm's argument. There is something extraordinary about the person of Socrates that the drug addict lacks. It is something by means of which we may praise Socrates and blame the drug addict. That is precisely Socrates's natural end as a person, which is his perseverance or *pervelle* to keep uprightness of will at all times and at all cost. It involves a permanent questioning of the desirability of his own desires, an inner monologue that leads him to the conclusion that a life dedicated to philosophy is the best thing he could pursue. This reflection on *the desirability of his own desires* is what the drug addict lacks. He no longer has rectitude of will, because he freely abandoned it. In short, someone who acts through uprightness of will for its own sake is someone who is a rational person.

That is far cry from the libertarian paternalist's idealized Econ, to say the least. To possess rectitude requires much more than a mere calculus of the means to one's ends. It involves a profound examination of those ends themselves and the appropriateness of the means, thereby deploying all of the cardinal virtues.

Within Anselm's volitional anthropology, there is no room for two separate selves that are at odds with one another. Anselm considered the will endowed with two affects, one directed towards justice and the other towards benefits. Yet both are jointly co-constitutive to achieve happiness, provided that justice as a second-order desire is attributed its proper place in one's value hierarchy. In this respect, Anselm speaks of the *diversas voluntates* or "inclinations of the will" that could be either directed towards itself or towards something else. And yet, they remain aspects of one and the same entity. In the first book of *De concordia*, Anselm writes:

> When, for instance, someone uses sword, tongue, or oratorical power, the sword or the tongue or the oratorical ability as such is the same thing when used rightly or wrongly. In the same way the will we use for willing (like the faculty of reason we use for reasoning) is the same thing as such whether we use it rightly or wrongly. (*De concordia* I.7)

Since we are gifted with just one unique self, it follows that many of the desires, emotions and other intentions actually need not be cognitive deficiencies at all. Virtue and wisdom for instance, like the will itself, are qualities of character concerned with choice and lie in an appropriate mean towards the end. With regard to the virtue of temperance, it is a mean concerning the correct attribution of bodily pleasures that hold between the extremes of shortage and overabundance. Practical wisdom assigns every pleasant activity its proper place according to the rule of a wise person who understands the appropriate place of these activities in the much larger enterprise of living a human life.

Libertarian paternalism by contrast fails to tell us anything meaningful about virtue and wisdom. As a child of modernity, it has no intelligible theory about the emotions and the role they play in the shaping of man's life. An emotion may have at least four distinct meanings (Mulligan 1998, 161–88). First, it may denote a drive or instinct such as hunger. Secondly, it may denote sensations that arouse pleasure or pain. A third meaning takes moods into account, such as anxiety or joy. And in the final sense, we may speak about the emotions as intentional acts. In the latter case, the object of one's desiring or sensing can be assessed independently of one's subjective feelings. It involves certain cognitive dispositions such as the ability to judge, to remember and to evaluate. One can feel regret for something that one has done or failed to do. In that case, the regret turns into a particular stance towards the intentional object which can be either correct or incorrect. True emotions are intentional acts and in no way mental states. Intentional acts have a particular logical relationship between their parts that transcend the boundaries of time. The act of proving a theorem results in a product, namely a proven theorem. Similarly, the act of love results in the correct understanding that the object of love is worthy of admiration. Once we take the intentional element into account it becomes manifest why libertarian paternalism's alleged distinction between means and ends is blurred. The virtuous man who properly assesses his objects of desire to be either correct or incorrect thereby also scrutinizes whether the means are efficient, appropriate, and/or morally salient. The relationship between means and ends is analytic, to borrow a Kantian notion. It is for this reason that the distinction between libertarian and classical, "hard" paternalism is merely one of gradation, not of substance.

Many of the activities that are discredited according to the reductionist criteria of libertarian paternalism may turn out to be virtuous and thus participate in rectitude for its own sake. Having a conversation with a friend over a glass of wine and a fine cigar might be "unhealthy" but definitely strengthens bonds. A mother taking an unpaid afternoon off in order to attend her child's performance at the school theatre may jeopardize her career prospects in the long run, but she definitely acts virtuously out of parental love. Is there any good reason why we should interfere with their choices? If willingly acting against the ever-greedy "Econ" is the price that has to be paid for leading a flourishing life, so be it.

References

Anselm of Canterbury. 1998. *The Major Works*. Edited by Brian Davies and G.R. Evans. New York: Oxford University Press.

van Dun, Frank. 1986. "Economics and the Limits of Value-Free Science." *Reason Papers* 11: 17–32.

Fuller, Lon Luvois. 1969 [1964]. *The Morality of Law*. New Haven: Yale University Press.

Hausman, Daniel M. and McPherson, Michael S. 2006 [1996]. *Economic Analysis, Moral Philosophy and Public Policy*. New York: Cambridge University Press.

Mulligan, Kevin. 1998. “From Appropriate Emotions to Values.” *The Monist* 81, no. 1: 161–88.

Nash-Marshall, Siobhan. 2008. “ Free Will, Evil, and Saint-Anselm.” *The Saint Anselm Journal* 5, no. 2: 1–23.

Putnam, Hilary. 2002. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press.

Rebonato, Riccardo. 2012. *Taking Liberties: A Critical Examination of Libertarian Paternalism*. Basingstoke: Palgrave Macmillan.

Rutten, Emanuel. 2018. *Het retorische weten*. Amsterdam: Leesmagazijn.

Seymour Fahmy, Melissa. 2018. “Kantian Perspectives on Paternalism.” In *The Routledge Handbook of the Philosophy of Paternalism*, edited by Kale Grill and Jason Hanna, 96–107. London: Routledge.

Thaler, Richard H., and Sunstein, Cass R. 2008. *Nudge: Improving Decisions about Health, Wealth and Happiness*. London: Penguin.

Weber Max. 1949. *The Methodology of the Social Sciences*. Translated by Edward A. Shils and Henry A. Finch. New York: Free Press.

Wolf, Susan. 2005. “Freedom Within Reason.” In *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, edited by James Stacey Taylor, 258–74. New York: Cambridge University Press.