The Fable of the Beekeepers

[Preliminary draft prepared for the 16th international conference of the Charles Gide Association for the Study of Economic Thought]

Guilhem Lecouteux

University of Bristol

Abstract (150 words)

The object of this paper is to question the widespread use in economics of the figure of the *social planner*. The conception of human societies for conventional economics is described as a 'hive' in which the individuals are producing welfare, scientifically designed by a social planner so as to maximise the total production of welfare. This perspective implicitly assumes that individuals are simply passive *loci* of experience, and denies their status of *agents* – i.e. that individuals can actively contribute to the shaping of their own preferences. I suggest that normative economists should leave the third person perspective of the social planner and ground their normative assessment on the second person standpoint (Darwall 2006a): the core of normative economics would then be individual autonomy and democratic processes, instead of the satisfaction of individual preferences. I illustrate this argument by discussing Ostrom's institutional design principles ensuring the sustainable management of common pool-resources.

Neoclassical welfare economics can be seen in the tradition of the utilitarian philosophy of the 18^{th} and 19^{th} century, since one of its main concerns is the question of whether a society composed of free and self-interested individuals can achieve the common good. The 'common good' is traditionally defined from the personal interests of the different individuals constituting the society, and a core theoretical issue of welfare economics is to define rules of aggregation of individual preferences so as to define social welfare functions. The question that follows is then to know whether the social order resulting from decentralised decisions constitutes a desirable state (according to this social welfare function), and in particular whether it can be outperformed by a constructed order build by an omniscient and omnipotent *social planner*. The concern for this philosophical question is clearly stated by the mere designation of the two '*fundamental* theorems of welfare economics', which precisely establish a theoretical connection between market equilibrium and Pareto optimality. Under certain institutional settings, it is however not certain that the first best will be spontaneously achieved: *market failures* occur when the social institution of competitive markets is not well-functioning, leading to a non-efficient outcome. It falls then to economists to produce policy recommendations consisting in changes in the institutional design that will *in fine* restore the efficiency of the market (such that internalising a negative externality with a tax). Those recommendations are then addressed to an entity indifferently called 'the government', the 'policy-maker', the 'social planner' or – within the context of libertarian paternalism (Sunstein and Thaler 2003) – the 'choice architect'. The social planner has therefore a dual role: on the one hand defining the social optimum (what could be achieved if all the decisions were centralised by an omniscient and omnipotent agent), and on the other hand designing the institutional environment of the society – so as to ensure the efficiency of the market equilibrium.

The aim of this paper is to highlight that an approach grounded on the figure of the social planner could offer a distorted analysis of many economic and social issues, since it neglects the role of democratic processes and reduces the society to a 'hive' that should be scientifically designed to increase its efficiency. I argue that normative economics build on a social planner assessment fundamentally deny individuals' autonomy and their ability to make normative assessments. This is due to a very specific perception of human societies, and to the tendency to think the world as if it were a logical construction – in which freedom has no place. I draw an analogy between the role of the social planner and a beekeeper: individuals are simply considered as passive ‘welfare producers’ and the role of the social planner is to improve the structure of the society in order to maximise social welfare, like a beekeeper who intends to maximise the production of honey by designing the structure of the beehive. I argue that endorsing the third person perspective of the social planner causes a fundamental ethical problem for normative economists, i.e. the definition of the normative criterion. I defend instead the idea that we should endorse a second-person standpoint (Darwall 2006) to produce normative prescriptions. Normative economists should not therefore try to define what is good for society by endorsing the social planner’s perspective (and in particular impose their own normative views about the society), but rather ensure the well-functioning of democratic processes, enabling the individuals to decide themselves what matters for their society.This position is relatively close to what Qizilbash (2011) labels the ‘thick view’of Sen’s capability approach (Sen 1999, 2009), according to which public reasoning plays a central role in the definition of the collective preferences of the society: it is argued that theoretical reasoning cannot, on its own, provide a normative criterion for individuals with different views (Sen 2004, 78). It therefore falls to the citizens, as a matter of sovereignty, to decide collectively what matters for themselves.

1. The market and the hive

Neoclassical economists see in the theoretical framework of competitive markets an efficient mechanism to allocate resources and satisfy individual preferences at the aggregate level. It has therefore been elevated to the status of an *ideal*: economists should use this theoretical framework to detect deviations from the ideal outcome, and then define the policies that should be implemented to correct those deviations, so as to reach *in fine* the socially desirable competitive market equilibrium. Four kinds of situations are then often considered as *market failures*, i.e. situations in which the market equilibrium differs from the Pareto efficient competitive equilibrium: externalities (Pigou 1920), market power, asymmetric information (Akerlof 1970) and bounded rationality[1] (Bennett et al. 2010, Sunstein 2014a). Economic analysis then provides a large toolbox to the social planner to fix those market failures and restore the first best. We can mention for instance the Pigovian tax to internalise negative externalities, Lindhal prices (Lindhal 1919) for the provision of public goods (which can be understood as a form of positive externality), preference revelation mechanisms (Vickrey 1961) in the case of asymmetric information, and nudges for boundedly rational individuals (Thaler and Sunstein 2008). The object of economics consists in designing different types of *incentives* (mainly monetary, but also psychological ones in the case of nudges) so as to overcome those market failures.

Grant (2002) suggests that the main reason why incentives seem to be ethically unproblematic (and therefore that they became central in economics) is that they can be assimilated to a form of *trade*. Consider an individual (P1) who intends to make an option A more attractive than the other options to another individual (P2). The incentive typically consists in a cash payment, such that the option A becomes unambiguously more interesting from P2's perspective (P2 may therefore decide to choose A, although he would have chosen an option B in the absence of incentive). P1 and P2 therefore reach an outcome that is beneficial for both of them, since P1 preferred that P2 chose A rather than B, and P2 has a strictly higher payoff when choosing A rather than B. An incentive therefore involves a voluntary action by all parties (P1 chose the level of the incentive, and P2 was not forced to choose the option A) leading to a mutually beneficial outcome: incentives are therefore ethically preferable to other forms of interventions such as coercion[2]. Grant then highlights that, although economists commonly believe that their thinking about political economy is in the continuation of the Scottish and English Enlightenment, the word 'incentive' does not appear in any of their writings[3]. Grant suggests that this evolution is closely related to the emergence of behavioural psychology in the early 20th century, and in particular to the

[1]It should be noted that, although bounded rationality can lead to another outcome than the competitive equilibrium, the view that this constitutes a market failure of the same nature than externalities is more controversial (see e.g. Lunn 2013).
[2]Following the same line of argument, it can also be argued that nudges are ethically unproblematic. Suppose that P1 is a benevolent choice architect who knows that P2 truly prefers A over B, but that P2 is likely to choose B due to some psychological bias. Here again, the nudge allows the players to reach a mutually beneficial outcome (since P1 is benevolent, she indeed only cares about P2's welfare), without forcing P2 to choose A.
[3]Grant simply mentions an anecdotal use by J.S. Mill in the *Principles of Political Economy*, and by Ricardo in the *Principles of Political Economy and Taxation.*

scientific management developed by Taylor (1911), that deeply influenced the way economists perceive the functioning of a market society. This evolution can be captured by a fundamental shift in metaphor from Smith's 'invisible hand' and the British thinkers of the 18th and 19th century to the current figure of the 'social engineer': while the society was conceived as a huge *clock,* functioning automatically and predictably according to natural laws, it progressively became 'an amalgam of forces in constant flux that can be directed to bring about progress' (Grant 2002, 117). Since there is no reason *a priori* for the spontaneous order to be socially optimal, the social planner ought to implement the adequate incentives so as to steer individual behaviours into the right direction.

Let us suggest an alternative metaphor for representing how welfare economists (both neoclassical and behavioural) conceive human society, a giant *hive* designed by a *beekeeper*. A society is composed of a group of individuals interacting with each other within specific institutional rules. The only aim of an individual is to satisfy her preferences, and all her actions are guided towards the satisfaction of her preferences. Those preferences – as soon as they are complete and transitive, as postulated in neoclassical welfare economics – can be represented by utility functions defined over the set of actions. While bees' only objective is to produce honey, the individuals' only objective is to produce utility. The production of honey is beneficial to the bees (as a source of food), and the production of utility is beneficial to the individuals (as a source of welfare[4]). Analogously to man-made beehives which are scientifically designed to increase the production of honey compared to the production of a natural beehive (allowing the share of the surplus between the bees for their own consumption and the beekeeper), it is possible to scientifically design the society such that the total production of utility is higher than at the initial natural order. Lastly, while the beekeeper can introduce frames with some honeycomb to ease the production of honey, the social planner can build markets to ease the processes of exchange and matching, and then the production of utility. A central element of this analogy is that individuals – in line with behaviourism – are *passive* entities, and are only responding to external *stimuli*, such as monetary incentives and nudges.

[4]Here I implicitly accept a preference satisfaction account of welfare: however I am not arguing that preference satisfaction *is* welfare, I am simply describing the metaphor of the hive and therefore endorse an implicit view in welfare economics that equates preference satisfaction and welfare. See Hausman (2012) on the differences between preference satisfaction and welfare.

The economist is therefore looking at the society as a beekeeper is looking at a hive: her objective is to determine the optimal structure for the hive such that the total production is maximised. Since individuals are perfectly responsive to incentives and nudges, their actions can be guided by the planner so as to compensate possible market imperfections. By only altering the institutional environment of the hive (such as implementing taxes – that do not fundamentally change the functioning of the market – or nudges), the individuals are not coerced, and can still refuse to choose the option the planner wants them to choose (although it is more expensive, you can for instance still buy a taxed good).

Economists, as social scientists, can then endorse this role of *social designer*, as clearly stated by Roth (1991):

> 'In the long term, the real test of our success will not be merely how well we understand the general principles which govern economic interactions, but how well we can bring this knowledge to bear on practical questions of microeconomic engineering, to design appropriate mechanisms for price formation [...], dispute resolution, executive compensation, market organisation, etc. [...] Just as chemical engineers *are called upon* not merely to understand the principles which govern chemical plants, but to design them, and just as physicians aim not merely to understand the biological causes of diseases, but its treatment and prevention, a measure of the success of microeconomics will be the extent to which it becomes the source of practical advice, solidly grounded in well tested theory, on designing the institutions through which we interact with one another' (Roth 1991, 113, my emphasis)

Neoclassical economists – thanks to the fundamental theorems of welfare economics – claim to have scientifically shown that competitive market is the best way to coordinate individual actions. The 'duty' of economists, as social scientists, is now to apply their results to real societies, by designing markets and price mechanisms that will rule the interactions between individuals. Those interactions are not restricted to standard economic interactions (such as buying and selling goods), but include 'some of the most important markets that we are involved in – the matching markets that determine what schools we go to, what jobs we get, and maybe who we are married to' (Roth 2012, 343). Human societies can then be reduced to a vast market in which welfare is produced through exchange and matching: the ideal society economists want to design is therefore a complete market (for which the first fundamental theorem holds). The society should therefore be nothing more than a purely mechanical device designed to maximise the production of welfare, in which nothing else than efficiency and preference satisfaction should matter. This point is particularly striking in Roth's discussion of *repugnance*. Repugnance concerns certain types of transaction (organ markets) or activities (dwarf tossing) that are considered as morally unacceptable. Roth (2007) however considers that repugnance is similar in nature to a 'difficult technological barrier', i.e. to a constraint that should eventually be overcome so as to maximise social welfare:

> The persistence of repugnance in many markets doesn't mean that economists should give up on the important educational role of pointing to inefficiencies and tradeoffs and costs and benefits. But neither should economists expect such arguments to immediately win every debate. Being aware of the sources of repugnance can only help make such discussions more productive, not least because it can help separate the issues that are fundamentally empirical -- like the degree of crowding out of altruistic donations that might result from different incentive schemes compared to how much new supply might be produced -- from areas of disagreement that are not primarily empirical. (Roth 2007)

Roth's claim is that empirical phenomena matter (crowding out effects for instance), but moral considerations and the disapproval of certain transactions on non-empirical basis (such as ethical or religious concerns) are not relevant. Economists should convince the population that considering a transaction as repugnant is pointless[5].

Economists see the world as if it was a giant hive, in which each individual seeks to maximise her individual welfare. They take the viewpoint of the beekeeper of this hive, who wants to maximise the total welfare produced within the hive, and who is also able to design the adequate incentives such that the individuals adopt the socially optimal behaviour.

[5]Note that in certain cases, the repugnance of certain activities seems to be incoherent with the acceptance of others, such as horse eating in California, which is forbidden for human consumption but not for pets (Roth 2007, 37). The cases of repugnance that Roth has in mind could however be much more problematic from an ethical perspective, such as kidney exchange.

While producing policy recommendations, they are therefore endorsing the beekeeper position (so as to define the optimal policies), and then address their recommendations to the same abstract beekeeper. The interpretation of those policy recommendations are therefore 'if I were an omnipotent and omniscient planner, this is what I would do' (Sugden 2013), or more precisely, 'if I were an omnipotent and omniscient planner, *and* if the real world was like the hive, *and* if what matters was preference satisfaction, then this is what I would do'.

The viewpoint from which normative assessments are made is the one of an impartially benevolent spectator (Sugden 2013): the economist builds a model in which idealised agents interact according to predefined rules (individuals are actuated by an intrinsic motive of preference satisfaction, and steered by different incentives), sets the model in motion, and then assesses the final outcome, from her position as a modeller. Therefore – unless her interests as an economist are directly related to the interests of a specific class of agents in her model (the ones funding her research for instance) – her assessment is impartial. Economists adopt a *third-person perspective* so as to provide normative assessments. One of the main problems of this perspective is the *arbitrariness* of the normative criterion.

2. Second person standpoint and the definition of the normative criterion

Standard (neoclassical but also behavioural) welfare economics adopts a third person perspective to provide normative assessments. Adopting such a neutral perspective – coined by Nagel (1986) as the 'view from nowhere' – requires the external third-person spectator to have an external criterion for her normative evaluation (Carrasco 2011), such as the quantity of happiness or the satisfaction of individual preferences. This is for instance on this basis that Conly (2013) argues that autonomy 'has been overvalued' (p.16), defining autonomy in the sense of Feinberg (1986), as 'the right to make choices and decisions – what to put in my body, what contacts with my body to permit, where and how to move my body though public space, how to use my chattels and personal property, what personal information to disclose to others, what information to conceal, and more' (p.54). Conly's central claim is indeed that letting people make their own choices may lead them to cause harm to themselves: there is therefore an *agent-neutral reason*[6] that justifies paternalism, the claim that satisfying the true preferences of any individual is normatively desirable. The satisfaction of one's true preferences is arbitrarily defined as the normative criterion for the third-person spectator, and therefore provides a reason for not respecting individual autonomy. Individual autonomy is however only considered in an instrumental perspective, as providing welfare *per se* (I may indeed prefer making my choices on my own), or as a necessary condition for one's self-development (which is beneficial *in fine* for all the society, since it for instance does not bound individual originality, which is necessary for society to progress).

Questioning the validity of autonomy as a means to satisfy one's true preferences however does not imply that a claim for autonomy is not normatively grounded: it is indeed possible to justify the claim for autonomy from a 'second-person standpoint' (Darwall 2006), defined as 'the perspective you and I take up when we make and acknowledge claims on one another's conduct and will' (p.3). The second-person standpoint requires being able to put oneself in another's shoes so as to simulate the reasoning of others and attribute them mental states (see e.g. Goldman (1989, 1992) and Gordon (1986, 1992)).Darwall (2006) for instance argues that Adam Smith's idea of sympathy in the *Theory of Moral Sentiments* makes him 'one of first philosophers of the "second person", if not the very first' (p.46): Carrasco (2011) then argues that Smith's impartial spectator is actually not in a third but a second-person perspective, and that 'when judging from inside the situation, the spectator perceives some qualities that are unreachable to the external observer, and one of them is, precisely, propriety'.Darwall (2006a, b) sees for instance in the second-person standpoint the roots of our moral responsibility, and then of the normativity of individual autonomy as part of respect for the dignity of persons: the autonomy of the will gives to the agents the ability to endorse a second-person standpoint, and then 'the authority, as a person, to make

[6]A reason is defined as *agent-neutral* (by opposition to an *agent-relative* reason) if it does not derive from a normative fact concerning an individual in particular. Nudging you to Increase your savings on the basis that it will maximise happiness is an agent-neutral reason: promoting welfare (even subjective welfare) is indeed a reason that can be stated for another individual than you. But nudging you on the basis that it will induce you to quit smoking (because you will not be able financially to continue buying cigarettes) is an agent-relative reason: the formulation of the reason indeed essentially refers to an individual and her addiction. On the differences between agent-neutral and agent-relative reasons, seeParfit (1984) and Nagel (1986).

claims and demands of one another as rational and free' (Darwall 2006, fn.11}. They have therefore the right to claim to be allowed to make their own choices: autonomy is therefore valuable for itself, and not valuable in an instrumental perspective – as suggested by Conly (2013) – as a means to promote welfare. The normative issue of paternalism, even if our welfare is not maximised, is therefore that it is 'a failure of respect, a failure to recognize the authority that persons have to demand, within certain limits, that they be allowed to make their own choices for themselves' (Darwall 2006, 268, see also Shiffrin (2000) for a similar argument).

By endorsing the social planner's perspective, normative economists must define a normative criterion to evaluate the outcomes of their models. It seems however problematic that the definition of such a criterion falls to economists rather than to citizens. On the contrary, acknowledging the possibility for the individuals to simulate the reasoning of others implies that we do not need to define an external normative criterion: what matters is indeed the autonomy of the agents (for a more precise definition of this notion of autonomy, see Lecouteux 2015, pp.129-134).

3. Normative economics and democracy

Focusing on the development of individual autonomy rather than preference satisfaction questions the standard view that normative economics is addressed to a social planner or choice architect (whose objective is to ensure the satisfaction of individual preferences), rather than directly to the individuals (McQuillin and Sugden 2012, 556).

3.1 The social planner and the Leviathan

By endorsing the social planner position, economists simultaneously endorse two roles: defining what is socially desirable, and how to achieve a socially desirable state. The Sovereign, as the holder of the supreme authority over society, is the only entity legitimate to define what matters for the society, while the Government, as the depositary of the Sovereign's authority to implement her will, is in charge of designing the society to achieve the Sovereign's will. Taking the viewpoint of the social planner implies that economists define the normative criterion of the society: but are they legitimate to replace the Sovereign in her normative assessments? I argue here that this position is justifiable within the context of the hive, when individuals have exogenous and coherent preferences, and is consistent with Hobbes contractarianism. However, as soon as we consider that individuals have some autonomy and can choose to some extent their own preferences, we should abandon this perspective: this will enable us to pass from a Hobbesian analysis of the social contract and of the role of the social planner to Rousseau's contractualism and the central place of democratic processes in the definition of individual preferences.

Suppose that the world can reasonably be represented by the hive described above, and that the individuals – as in conventional welfare economics – are actuated by the satisfaction of exogenously given preferences. Although the first fundamental theorem of welfare economics ensures that the free interaction of the individuals leads to a socially desirable outcome in competitive markets, they are not able to reach a Pareto efficient outcome if they face collective action situations or interact in an imperfect market. This suboptimal situation is quite similar to Hobbes'*state of nature*, and his idea of a 'war of all against all': there does not exist any entity at the collective level to ensure the pacific coexistence of all the individuals, whose interests are generally conflicting (they have in particular no reason to trust each other, since they cannot be sure that the others will respect their promises). The hive is therefore in a suboptimal state of anarchy, and it would be in the interest of each individual to define an entity at the aggregate level such that the satisfaction of its collective preferences would lead to a Pareto-superior outcome. Such an entity would be the *Sovereign* of the society, since its preferences would define what is preferred by the society as a whole. But how can we define the collective preferences of the Sovereign? Or put differently, how can we define what matters at the collective level, given the preferences of each individual (whose satisfaction matters)? The difficulty of this question is that Arrow's impossibility theorem precisely states that such collective preferences cannot exist if we want to verify some basic and intuitive axioms: non-dictatorship, universality, independence of irrelevant alternatives and unanimity (Arrow 1951).

The first solution suggested so as to 'save' the hive from this suboptimal state of anarchy is therefore the advent of a dictator. This is for instance Hobbes' position, who considers that men are fundamentally unable to coexist peacefully without a supreme authority. Hobbes conceptualised the social contract in those terms:

> 'This is more than consent, or concord; it is a real unity of them all in one and the same person, made by covenant of every man with every man, in such manner as if every man should say to every man: I authorise and give up my right of governing myself to this man, or to this assembly of men, on this condition; that thou give up, thy right to him, and authorise all his actions in like manner' (Hobbes 1651, 106)

The individuals therefore voluntarily give up their 'right to govern [themselves]' to another individual or group of individuals, on the condition that everyone does the same. In this situation, the collective preferences are defined exclusively by the preferences of a single individual (or of a subgroup of the population). The different individuals must then respect the will of the Sovereign: although the Sovereign's preferences may not be consistent with their own preferences, this situation is better than the state of anarchy without Sovereign (they achieve here a second best). This may be why the individuals merely*authorise* the Sovereign to govern them: the power of the Sovereign over the rest of the individuals is not absolute, since it must provide them more than what they would have gotten in the state of nature (safety and peace according to Hobbes). In a standard microeconomic context, if the market equilibrium is not Pareto efficient, the individuals authorise the social planner to implement public policies such that the new equilibrium is a Pareto improvement. They have however nothing to say about the way the social planner shares the surplus between the individuals: the collective preferences of the social planner are indeed arbitrarily defined by the modeller, in general as a sum of individual utilities.

An alternative solution is grounded on the observation that in competitive markets, the initial situation of anarchy within the Hive is actually not suboptimal: it seems therefore possible to ensure the peaceful coexistence of the individuals, and this without coercing their freedom by designating a dictator. It is likely that Hobbes did not think of this solution because he believed that the interaction of self-interested individuals could only lead to an open conflict: Gauthier (1969) indeed argues that 'it is impossible to emphasize too strongly that it is the substantive premises about human nature, and not the formal structure of the theory, that determines its absolutist character'. If self-interested individuals can coexist and reach a desirable state in competitive markets, then a solution to improve the society while avoiding the appeal to a dictator is to build markets, i.e. to ensure that all the interactions in society are ruled by competitive markets. This is precisely the approach defended by neoclassical economics.

However, just as Gauthier argues that the absolutist character of Hobbes theory is due to its anthropological premises, we can argue that this reasoning is valid if and only if the underlying anthropological model of the Hive makes sense. In particular, humans should be passive actors fundamentally guided by the satisfaction of their preferences. The reason why social cooperation – although it is clearly in the interest of each individual – cannot be naturally reached is that the individuals described in the hive are not able to form commitments. They cannot decide to go beyond their immediate interest, since they are programmed to systematically choose the action that satisfies their best interest. The Leviathan – through its absolutist character – confers them indirectly this power of commitment[7]. Alternatively, competitive markets can replace the social dimension of exchange (that could generate some conflict) with a disembodied price system: a perfectly competitive market would be a 'morally free zone, a zone in which the constraints of morality would have no place' (Gauthier 1986, 84). Indeed, according to Gauthier, since the common good is an unintended by-product of individuals' pursuit of their self-interest in competitive markets, market relationships do not need to be genuinely social (Bruni and Sugden 2008, 38).

Assuming that individual preferences are exogenously given seems therefore to legitimate the standard approach in normative economics, according to which we should take the standpoint of a benevolent despot so as to provide normative assessments. The normative criterion may not be in the best interest of each player (such as for instance the

[7]Hobbes's position is that cooperation among large groups is impossible in the absence of a supreme political authority, since the threat of sanctions is necessary to give the insurance to everyone that the others will cooperate: it is not because not cooperating is too costly that cooperation is ensured, but because it is actually rational for each individual to cooperate when they are ensured that all the others are committed to reciprocate (Hollis 1998).

maximisation of the sum of utilities, since each player would individually prefer to have a higher weight than the others), but they should however be ensured to get at least the level of utility they would have achieved at the market equilibrium without intervention. Neoclassical economists can therefore defend their approach on the basis that it provides a contractarian solution to the initial suboptimal equilibrium: the existence of the social planner (and therefore the legitimacy for economists to endorse this position) is justified because it enables the individuals to reach a mutually beneficial outcome, compared to the initial situation without intervention. This approach however does not respect individuals' autonomy and their right to decide by themselves the terms of the agreement (we can indeed notice that the choice of the normative criterion is left to the economist's discretion, without any control of the individuals).

3.2 Autonomy and the social contract

Suppose now that, unlike within the model of the hive, individual preferences are not necessarily fixed, and players can make commitments. We can now provide two other solutions to improve the initial equilibrium. The first solution is that rational individuals may decide to make the commitment of not breaking agreements, by grounding principles of morality on contractarianism (Gauthier 1986, 1991). Unlike Hobbes who suggested that the existence of the Leviathan was required to provide the insurance to all the individuals that the other individuals would be committed to cooperate, Gauthier argues that rational individuals have the ability to choose to follow principles of morality that go against their direct self-interest, but that allow them to reach mutually beneficial agreements with other 'moral' players. Gauthier indeed suggests that rational players may either be *straightforward maximisers*, who always choose what is in their best interest (and who may therefore break non-binding agreements such as 'cooperate in a prisoner's dilemma when the other cooperates'), or *constrained maximisers*, who make the commitment of not breaking agreements. Gauthier then argues that, in a world in which the type of maximiser (constrained or straightforward) of each individual is perfectly observed, then constrained maximisers will outperform straightforward maximisers[8]. In this situation, the individuals do not need a benevolent despot to enforce their commitments: it is indeed in their own interest, as rational agents, to choose to become constrained maximisers, because it throws the foundations of reciprocity within the society.

The second solution is slightly different from Hobbes and Gauthier's contractarianism, since it offers a contractualist solution to the initial suboptimal equilibrium. The distinction between contractarianism and contractualism is that contractarianism takes moral principles to result from rationally self-interested bargaining, while contractualism sees the agreement as governed by a moral ideal of equal respect (Darwall 2003, 4). Contractarians indeed assume that they have a natural right to claim the outcome they would achieve in the absence of agreement: the agreement is then build as a mutually beneficial agreement on the basis of this disagreement outcome. Contractualists, on the other hand, consider that this moral claim on the disagreement outcome is arbitrary, and therefore that the moral principles resulting from the bargaining on this initial outcome have no moral force. Contractualists are not trying to gain acceptance from rules they prescribe from the perspective of their own interest, but they are prescribing and agreeing on rules from a common perspective as one free and equal person among others. This is typically Rousseau's position and his conception of the political community of citizens, as a form of association in which each, 'uniting with all, nevertheless obeys only himself'. The Kantian categorical imperative rests on a similar logic, since the moral law the individual prescribes to herself should become a universal law. A similar approach is developed by Rawls (1971), since the original position, under the veil of ignorance, implies that the individuals are disconnected from their personal interests (although they are deliberating so as to satisfy their individual interest, they have no idea of what will be their position in the society). The principles of justice they are choosing are therefore chosen among a community of free and equal individuals. More recently,Parfit (2010) suggests the*Kantian Contractualist Formula* as an explicitly contractualist reformulation of Kant's categorical imperative, according to which 'everyone ought to follow the principles whose universal acceptance everyone could rationally will' (p.20).

[8] See Lecouteux 2015, 226, for a formal proof.

An interesting difference between the contractarian and the contractualist approaches is that, while the notion of*authorisation* is central with Hobbes, contractualists like Rousseau put a strong emphasis on the notion of *commitment*:

> "Each of us puts his person and all his power in common under the supreme direction of the general will, and, in our corporate capacity, we receive each member as an indivisible part of the whole. [...] This formula shows us that the act of association comprises a mutual undertaking between the public and the individuals, and that each individual, in making a contract, as we may say, with himself, is bound in a double capacity; as a member of the Sovereign he is bound to the individuals, and as a member of the State to the Sovereign." (Rousseau 1762)

Rousseau's social contract gives rise to an emerging entity, the *People*, who is sovereign. Contrary to Hobbes who designated the Sovereign among the population – i.e. designated a dictator – Rousseau considers an emerging entity, the group as a distinct agent. The People has his own will, the *general will*, which can only be revealed by the enlightened deliberation of the citizens. Unlike the contractarian approaches in which the object of the contract can be seen as the fair aggregation of individual preferences, so as to reach a mutually beneficial outcome, the social contract consists in a transfer of agency from the individual to the collective level through a collective act of commitment (Hollis 1998): the individuals can then choose to become an 'indivisible part of the whole', so as to be actuated by the collective objective of the group rather than by their own interest.

Although the contractarian and the contractualist approaches differ in their motivations (contractarians are indeed ultimately self-interested, while contractualists are interested in their community), they both induce a transfer of agency from the individual to the collective level: the individuals are looking for a *mutual advantage*, instead of trying to satisfy their own preferences. Social interactions, unlike within the hive where the only motive of action is the satisfaction of one's own preferences, can be understood as joint intentions for mutual assistance (Bruni and Sugden 2008). Individuals looking for mutual advantage rather than their own self-interest can then engage in *team reasoning* (Sugden 1993, Bacharach 2006), i.e. choose their action by considering themselves as a part of a larger entity, the group of individuals within which the agreement is recognised. Although team reasoning is an end in itself for a contractualist, it is only instrumental for the contractarian, as a means to achieve *in fine* her own interest.

On a more practical level, it is not sure whether the individuals are fully aware of the objective of the others: they must then discuss to identify what outcome could be mutually beneficial (this is not necessary within the hive, since the individuals do not care about the objective of others – they are only interested in the satisfaction of their own preferences). It is only after this phase of identification of the interest of each individual that a consensus may emerge concerning the possible mutually beneficial outcomes. The viewpoint from which normative assessments are made is therefore not a 'view from nowhere', outside society, but a 'view from everywhere', everywhere within the society. Reasons for actions are then agent-relative rather than agent-neutral: normative claims are indeed directly related to the actual agents of the group, and not to a normative principle stated by a third-person observer.

Neoclassical and behavioural welfare economists, by trying to assess society from an outside position, may *in fine* impose their own normative views to other individuals – they are indeed not able to view the society from this third-person perspective. The only solution to produce normative assessments is to confront personal views within the society: normative economists, rather than trying to guess what matters for the individuals, should try to organise society such that the citizens can deliberate, and then form their own normative claims. Economists are not legitimate to decide what matters for the individuals: they should instead help the citizens to debate so as to identify mutually beneficial agreements. Their role is therefore not to replace the Sovereign and to advise the Government, but to directly advise the Sovereign on the mutually beneficial outcomes that may be reached.

The distinction we draw between neoclassical welfare economics -- grounded on a social planner assessment -- and our procedural form of normative economics is perfectly illustrated with the dual interpretation we can have of the capability approach (Qizilbash 2011, Baujard and Gilardone 2015). While Sen (2009) emphasises that what fundamentally matters is public debate and democracy, other authors (e.g. Nussbaum 2000) place themselves in the position of an expert and want to define a universal list of capabilities that could characterise a 'good life'. The latter – considered today as 'the' capability approach – is therefore endorsing a social planner perspective, aiming at defining what matters for the individuals, whereas the former considers that providing basic capabilities is not an objective in itself, but a necessary condition for the individuals to be able to make public debates.

4. The management of common pool resources

I now illustrate this claim by considering, on a more empirical level, the design principles that characterize robust institutions for managing common pool resources (Ostrom1990). The objective here is to highlight that the sustainable management of CPR is possible if and only if public policies are designed such that they provide to the individuals the means to organise themselves, rather than imposing from the outside a solution designed for correcting a market failure by the introduction of adequate incentives.

CPR are a class of goods characterized by two attributes, the difficulty of excluding individuals from benefiting from the resource, and the subtractability of the benefits consumed by an individual from those available to others. Two main types of problems can emerge in this context, appropriation and provision problems: appropriation problems are related to the exclusion of potential beneficiaries and the repartition of the output, whereas provision problems are related to the management of the stock of the resource, whether it be its creation, the maintenance or improvement of its production capabilities, or the avoidance of its destruction (Ostrom*et al.* 1994, 9). Ostrom (1990) suggests a list of eight design principles that characterize the institutions enabling a sustainable management of CPR, which have been slightly amended by Cox *et al.* (2010), who provide a meta-analysis of the different empirical works that tested those principles (extract from Cox*et al.* 2010, table 4):

- 1A: user boundaries: clear boundaries between legitimate users and non-users must be clearly defined;
- 1B: resource boundaries: clear boundaries are present that define a resource system and separate it from the larger biophysical environment;

- 2A: congruence with local conditions: appropriation and provision rules are congruent with local social and environmental conditions;
- 2B: appropriation and provision: the benefits obtained by users from a CPR, as determined by appropriation rules, are proportional to the amount of inputs required in the form of labour, material, or money, as determined by the provision rules;
- 3:collective-choice arrangements: most individuals affected by the operational rules can participate in modifying the operational rules;
- 4A: monitoring users: monitors who are accountable to the users monitor the appropriation and provision levels of the users;
- 4B: monitoring the resource: monitors who are accountable to the users monitor the conditions of the resource;
- 5: graduated sanctions: appropriators who violate operational rules are likely to be assessed graduated sanctions (depending on the seriousness and the context of the offense) by other appropriators, by officials accountable to the appropriators, or by both;
- 6: conflict-resolution mechanisms: appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials;
- 7: minimal recognition of rights to organize: the rights of appropriators to devise their own institutions are not challenged by external governmental authorities;
- 8: nested enterprises: appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises

My purpose is not to discuss extensively these different principles, but to highlight that most of them are directly supported by the conception of normative economics put forward in this paper as developing democratic processes rather than implementing incentives. We can indeed notice that the main feature of those principles is the idea that the users of the CPR should be able to design their own institutional environment (this is quite explicit in the principles 3 and 7). Furthermore, the possible external actors who monitor the users and the resource, or who assess possible sanctions in case of non-respect of the appropriation and provision rules are systematically accountable to the users (who therefore remain sovereign). Several empirical studies showed for instance that when the rules are imposed by an external authority, this one generally fails to enforce them, leading to suboptimal results (Ostrom*et al.*, 1994, 221-222). Nevertheless, although direct interventions often fail, the government can help the users to manage more efficiently the resource: Blomquist (1994)– from empirical evidence of groundwater systems in Southern California – suggests for instance that the design of provision and appropriation rules is facilitated by the presence of government agencies that can provide reliable information to the users (296-297). From various laboratory experiments and field studies, Ostrom*et al.* (1994) argue that the individuals can overcome the temptation of overusing the resource if they have some expectation of mutual trust, or the possibility of building trust through continued interaction and communication (328), and if they have some autonomy to decide on their own rules (323). However, since it appears that boundedly rational individuals can have some difficulties to reach optimal rules -- mainly due to information issues and the complexity of the problem -- governmental agencies play an important role by recognizing the right to the individuals to form their own rules and commitments, but also by providing them reliable information and backup enforcement mechanisms (322-327).

Those conditions, and in particular the role of the government as an actor who provides information and support to the individuals without directly intervening nor trying to influence individuals' choices, correspond to the kind of normative prescriptions that would result from the conception of normative economics defended in this paper. My claim is indeed that economists should assess public policies in terms of individual autonomy, i.e. the ability of the individuals to engage in public debate as informed citizens, so as to be able to collectively choose their own preferences: this requires providing the largest information to the individuals, and let them decide on their own rules rather than imposing external rules.

In addition, the case of CPR gives us another argument in favour of a more deontological formulation of normative economics, the impact of institutional rules on individual preferences. It seems indeed that individual preferences in CPR situations depend on the institutional organisation that rules the appropriation and the provision of the resource: self-organized institutions are more likely to generate prosocial behaviours than rules imposed by an external authority. It means that imposing the same policy can have a different impact according to its initiator: empirical evidence in CPR situations suggest that policies democratically implemented are more likely to be efficient than policies implemented by an external authority. It is therefore probably not equivalent to try to implement what the individuals would have chosen if they were autonomous (such as within a social planner's perspective) and to try to directly improve the autonomy of the individuals. A measure implemented by the individuals who will be directly affected by it may be more efficient than the same policy implemented by an external authority: in the latter case, the individuals can indeed be suspicious about the objective of the government, and then be affected by crowding-out effects (they may for instance be tempted to cheat and exploit the weaknesses of a system of monetary incentives implemented by an external authority).

The management of CPR offers a good illustration of one of the main objectives of the reformulation of normative economics I defend. While neoclassical welfare economists ground their normative assessments on consequentialist considerations such as the welfare generated by the satisfaction of one's preferences, I suggest adopting a more procedural approach by grounding our normative assessments on individual autonomy and the ability for the individuals to choose themselves what matters for them. It seems indeed that the sustainable management of a CPR (and therefore the welfare it generates) is not only the result of the implementation of specific rules, but also of the conditions under which those rules were decided: promoting individual welfare therefore requires promoting individual autonomy, since the rules that will enable the individuals to maximise their welfare are more likely to be efficient if they are implemented by autonomous agents rather than by an external authority.

References

Akerlof, G. (1970). “The Market for `Lemons': Quality Uncertainty and the Market Mechanism".*Quaterly Journal of Economics*, 84: 488-500.

Arrow, K. (1951). *Social Choice and Individual Values*. Yale University Press.

Bacharach, M. (2006). *Beyond Individual Choice. Teams and Frames in Game Theory*. Princeton University Press. Edited by Natalie Gold and Robert Sugden.

Baujard, A. and M. Gilardone. "Sen is not a capability theorist." *Available at SSRN 2589510* (2015).

Bennett, M., Fingleton, J., Fletcher, A., Hurley, L., and Ruck, D. (2010). “What Does BehaviouralEconomics Mean for Competition Policy?".*Competition Policy International*, 6: 111-137.

Blomquist, W. 1994. “Changing Rules, Changing Games: Evidence from Groundwater Systems in Southern California”. In *Rules, Games and Common-Pool Resources*, Ostrom, E., R. Gardner, and J. Walker. 293-300. Ann Arbor: University of Michigan Press.

Bruni, L. and Sugden, R. (2008). “Fraternity: Why the Market Need Not Be a Morally FreeZone". *Economics and Philosophy*, 24(1):35-64.

Carrasco, M. (2011). “Hutcheson, Smith and Utilitarianism". *The Review of Metaphysics*,64(3):515-553.

Conly, S. (2013). iAgainst Autonomy. Justifying Coercive Paternalism. Cambridge: CambridgeUniversity Press.

Cox, M., G. Arnold, and S. Villamayor Tomas. 2010. “A Review of Design Principles for Community-based Natural Resource Management”. *Ecology and Society*. 15(4): 38.

Darwall, S. (2003). *Contractarianism, Contractualism*. Wiley-Blackwell.

Darwall, S. (2006a). *The Second-Person Standpoint: Morality, Respect, and Accountability*.Harvard University Press.

Darwall, S. (2006b). “The Value of Autonomy and the Autonomy of the Will". *Ethics*, 116: 263 284.

Dworkin, G. (1972). “Paternalism". *Monist*, 56:64-84.

Feinberg, J. (1971). “Legal Paternalism". *Canadian Journal of Philosophy*, 1:106-124.

Gauthier, D. (1969). *The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes*.Oxford: Clarendon Press.

Gauthier, D. (1986). *Morals by Agreement*. Oxford: Oxford University Press.

Gauthier, D. (1991). “Why Contractarianism?". In Vallentyne, P., editor, *Contractarianism andRational Choice*, pages 15-30. Cambridge: Cambridge University Press.

Goldman, A. (1989). “Interpretation Psychologized". *Mind and Language*, 4:161-189.

Goldman, A. (1992). “In Defense of the Simulation Theory". *Mind and Language*, 7:104-119.

Gordon, R. (1986). “Folk Psychology as Simulation". *Mind and Language*, 1:158-171.

Gordon, R. (1992). “The Simulation Theory: Objections and Misconceptions". *Mind and Language*, 7:11-34.

Grant, R. (2002). “The Ethics of Incentives: Historical Origins and Contemporary Understandings".Economics and Philosophy, 18(1): 111-139.

Gutmann, A. (1987). *Democratic Education.* Princeton: Princeton University Press.

Hausman, D. (2012). *Preference, Value, Choice, and Welfare*. Cambridge University Press.

Hobbes, T. (1651). *Leviathan, or The Matter, Forme and Power of a Common Wealth Ecclesiasticall and Civill*. McMaster University Archive of the History of Economic Thought.

Hollis, M. (1998). *Trust within Reason*. Cambridge: Cambridge University Press.

Kant, I. (1797). *Metaphysics of Morals*. Cambridge: Cambridge Univsity Press.

Lecouteux, G. (2015). *Reconciling Normative and Behavioural Economics*. PhD thesis.

Lunn, P. 2013. “Are Consumer Decision-Making Phenomena a Fourth Market Failure? *ESRI Working Paper 455*.

McQuillin, B. and Sugden, R. (2012). “Reconciling Normative and Behavioural Economics: theProblems to be Solved". Social Choice and Welfare, 38:553-567.

Mill, J. (1859). *On Liberty*. Meridian Book. inUtilitarianism, On Liberty, Essay on Bentham,together with seleceted writings of Jeremy Bentham and John Austin (1974).

Nagel, T. (1986). *The View from Nowhere*. Oxford: Oxford University Press.

Nussbaum, M. (2000). *Women and Human Development*. Cambridge: Cambridge UniversityPress.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* New-York: Cambridge University Press.

Ostrom, E., R. Gardner, and J. Walker. 1994. *Rules, Games and Common-Pool Resources*. Ann Arbor: University of Michigan Press.

Parfit, D. (2010). *On What Matters*. Oxford University Press.

Pigou, A. (1920).*The Economics of Welfare*. London: McMillan.

Qizilbash, M. (2011). “Sugden's Critique of Sen's Capability Approach and the dangers oflibertarian paternalism". International Review of Economics, 58(1):21-42.

Rawls, J. (1971). A Theory of Justice. Cambridge: Harvard University Press.

Roth, A. (1991). “Game Theory as an art of Empirical Economics". The Economic Journal.

Roth, A. (2007). “Repugnance as a Constraint on Markets". Journal of Economic Perspectives,21(3): 37-58.

Roth, A. (2012). “The Theory and Practice of Market Design". In The Nobel Prize, pages 343-363.

Rousseau, J.-J. (1762). *Du Contrat Social*. Paris: LGF. Reprint 1996.

Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.

Sen, A. (2004). “Capabilities, Lists and Public Reason: Continuing the Conversation". *Feminist Economics*, 10(3):77-80.

Sen, A. (2009). *The Idea of Justice*. Cambridge MA: The Belknap Press of Harvard UniversityPress.

Shiffrin, S. (2000). “Paternalism, Unconscionability Doctrine, and Accommodation". Philosophy and Public Affairs, 29: 205-250.

Smith, A. (1759). *The Theory of Moral Sentiments*. Penguin. Reprint 2010.

Sugden, R. (1993). “Thinking as a Team: Toward an Explanation of NonselfishBehavior". *SocialPhilosophy and Policy*, 10: 69-89.

Sugden, R. (2008). “Why Incoherent Preferences do not Justify Paternalism*". ConstitutionalPolitical Economy*, 19(3):226-248.

Sugden, R. (2013). “The Behavioural Economist and the Social Planner: to Whom ShouldBehavioural Welfare Economics be Addressed?". *Inquiry*, 56:519-538.

Sunstein, C. (2014). *Why Nudge? The Politics of Libertarian Paternalism*. Yale UniversityPress.

Sunstein, C. and Thaler, R. (2003). “Libertarian Paternalism Is Not an Oxymoron". *TheUniversity of Chicago Law Review*, 70(4):1159-1202.

Taylor, F. (1911). *The Principles of Scientific Management and Shop Management*. Routledge/Thoemmes Press. Reprint 1993.

Thaler, R. and Sunstein, C. (2008). *Nudge. Improving Decisions About Health, Wealth, andHappiness*. Yale University Press.

Vickrey, W. (1961). \Counterspeculation, Auctions, and Competitive Sealed Tenders". *Journalof Finance*, 16(1): 8-37.