Neo-Samuelsonian Welfare Economics: From Economic to Normative Agency

Cyril Hédoin[*]

University of Reims Champagne-Ardenne, France

Version 1.0: 28/05/2017

Abstract: This paper explores possible foundations and directions for "Neo-Samuelsonian Welfare Economics" (NSWE). I argue that neo-Samuelsonian economics entails a reconciliation problem between positive and normative economics due to the fact that it cuts the relationship between economic agency (i.e. what and who the economic agent is) and normative agency (i.e. what should be the locus of welfare analysis). Developing a NSWE thus implies to find a way to articulate economic and normative agency. I explore two possibilities and argue that both are attractive but have radically different implications for the status of normative economics. The first possibility consists in fully endorsing a normative approach in terms of "formal welfarism" which is completely neutral regarding both the locus and the unit measure of welfare analysis. The main implication is then to make welfare economics a branch of *positive* economics. The second possibility is to consider that human persons should be regarded as axiologically relevant because while they are not prototypical economic agents, they have the ability to represent them both to themselves and to others as reasonable and reliable beings through narrative construction processes. This gives a justification for viewing well-being as being constituted by the persons' preferences, but only because these preferences are grounded on reasons and values defining the identity of the persons. This view is somehow compatible with recent accounts of well-being in terms of value-based life satisfaction and implies a sensible reconsideration of the foundations of welfare economics.

Keywords: Neo-Samuelsonian economics – Welfare Economics – Revealed preference theory – Preference-satisfaction view of welfare – Economic agency

1. Introduction

The nature and the scope of economics has recently been the subject of controversies, especially due to the rise of behavioral economics and neuroeconomics (e.g. Caplin and Schotter 2010). The growing importance of experimental methods in economics, added to the empirical evidence that individuals' behavior most of the time fails to satisfy the various rationality

[*] Professor of Economics, economics and management research center REGARDS (EA 6292). Contact: cyril.hedoin@univ-reims.fr

requirements traditionally set by economists, have contributed to feed the notion that the fate of economics and psychology is to get closer together. While this idea is vehemently resisted by some mainstream economists like Gul and Pesendorfer (2010) or Binmore (2009), it is on the ground of an endorsement of revealed preference theory that many philosophers of economics like Hausman (2011), following Sen's (1973) early critique, find unappealing. Through a set of writings, Don Ross [e.g. (2005); (2011); (2014a); (2014b)] has been engaged in an attempt to give more solid philosophical and theoretical foundations to what he labels "Neo-Samuelsonian Economics" (henceforth, NSE). NSE is the continuation of a long tradition in economics including the works of Pareto, Hicks, Samuelson, Savage and others, responsible for the separation of economics and psychology in the first half of the 20th century. To eschew any form of behaviorism, Ross argues that NSE should be grounded on a sophisticated philosophy of mind account of mental states building on functionalism and semantic externalism. Daniel Dennett's (1989) "intentional stance functionalism" in particular plays a pivotal role in Ross's account.

In an attempt to assess Sen's critique of revealed preference theory viewed through the lens of NSE, Hédoin (2016) argues that NSE may be seen as a successful rebuttal of Sen's critique as far as positive economics is concerned. However, he also suggests that the NSE defense of revealed preference theory comes at a cost: it renders the articulation between descriptive and explanatory accounts of market dynamics and the evaluation of market outcomes (i.e. normative economics) unclear at best. The reason is that NSE promulgates the idea that human persons are not prototypical economic agents. However, mainstream welfare economics has been historically grounded since Pareto's days on the notion of "consumer sovereignty" according to which welfare assessment should be conducted with respect to the satisfaction of *persons' preferences* (McQuillin and Sugden 2012). The consumer sovereignty principle has itself be defended on the basis of the two following postulates: (i) individuals are rational (i.e. have consistent preferences according to a set of axioms) and (ii) their preferences are constitutive of their well-being. Together, these two postulates provide a way to bridge positive and normative economics in a welfarist perspective. The starting point of this paper is the claim that NSE annihilates postulate (i) and makes postulate (ii) gratuitous. Ironically, though for very different reasons, NSE and behavioral economics thus both lead to what McQuillin and Sugden (2012) have characterized as a "reconciliation problem" between positive and normative economics.

My main aim in this paper is to explore possible foundations and directions for what I will call "Neo-Samuelsonian Welfare Economics" (henceforth, NSWE). I argue that the reconciliation problem comes from the fact that NSE cuts the relationship between economic agency (i.e. what and who the economic agent is) and normative agency (i.e. what should be the locus of welfare analysis). Developing a NSWE thus implies to find a way to articulate economic and normative agency. I explore two possibilities and argue that both are attractive but have radically different implications for the status of normative economics. The first possibility consists in fully endorsing a normative approach in terms of "formal welfarism" (Fleurbaey 2003) which is completely neutral regarding both the locus and the unit measure of welfare analysis. Interestingly, this is essentially the route taken by the so-called "behavioral welfare economics". The main implication is then to make welfare economics a branch of *positive* economics. The second possibility is more original but somehow coherent with Ross's own views: human persons should be regarded as axiologically relevant because while they are not prototypical economic agents, they have the ability to represent them both to themselves and to others as reasonable and reliable beings through narrative construction processes. This gives a justification for viewing well-being as being constituted by the persons' preferences, but only because these preferences are grounded on reasons and values defining the identity of the persons. This view is somehow compatible with recent accounts of well-being in terms of value-based life satisfaction and implies a sensible reconsideration of the foundations of welfare economics.

The rest of the paper is organized as follows. Section 2 puts NSE in perspective with contemporary developments in revealed preference theory. Section 3 describes the account of economic agency underlying NSE. Section 4 explains why NSE leads to a reconciliation problem between positive and normative economics and points out that the reasons are significantly different than for behavioral economics. Section 5 briefly surveys recent experimental works taking the stance of NSE to make welfare assessments. It suggests that they are ambivalent regarding the articulation between positive and normative economics. Section 6 explores a possible reconciliation based on formal welfarism. Section 7 deals with a second possible reconciliation grounded on the narrative abilities of human persons. Section 8 briefly concludes.

2. Contemporary Revealed Preference Theory and Neo-Samuelsonian Economics

The roots of NSE can be found in the early 20th century at the very beginning of what has been later known in textbooks as "neoclassical economics". As it is well-known, early marginalist economists such as Stanley Jevons and Francis Edgeworth developed a theory of markets and exchanges building on a theory of cardinal utility largely inspired from Bentham's utilitarianism. Edgeworth's notion of "hedonimeter" and the fact that he cites Weber's and Fechner's work on "psychophysics" are indeed good indications that the first marginalists essentially viewed utility as a psychological and scalar measure.[1] The ordinalist revolution launched by Pareto and Hicks led to a progressive emancipation of economics from psychology as it was demonstrated that consumer theory can be entirely reformulated without assuming that utility is a cardinal and psychological measure. In an ordinalist perspective, utility merely corresponds to the mathematical representation of a preference ordering over bundle of goods. Paul Samuelson's (1938) and others' work within what is nowadays known as revealed-preference theory definitely separated economics from psychology: Samuelson and his followers indeed shown that under plausible conditions the axioms and results of ordinal consumer theory can be expressed in purely behavioral terms, i.e. choice behavior. In particular, it has been demonstrated that under such conditions, the various axioms of revealed-preference theory (WARP, SARP, GARP) are equivalent to the maximization of some set of utility functions.

Psychology and economics have since essentially remained separated, largely due to revealed-preference theorists' efforts. However, the attempt to rebuild microeconomic theory on purely behavioral foundations has been attacked very early on, most famously by Sen [(1973); (1977)]. Sen's critique articulates two related points: (i) the rationality axioms on which choice theory and microeconomics relied are plausible only if the foundational concepts (preferences, utility)

[1] On Edgeworth's hedonimeter, see Colander (2007).

are understood in a "psychological" or "mentalist" sense;[2] (ii) (micro)economic theory should not conflate choices, personal preferences and welfare (Hédoin 2016). In spite of Sen's and others' similar critiques, the revealed preference approach remains officially (as one can judge by looking at leading microeconomic textbooks) the doctrine on which virtually all mainstream economics builds.[3] The emergence of the behavioral economics and neuroeconomics research programs has however started to shift the pendulum within the economic profession. On the one hand, the development of behavioral economics and its use of experiments have encouraged a focus on the properties of individual behavior and its underlying psychology (loss aversion, endowment effect, framing effect and so on). On the other hand, the increasing use of techniques coming from neurosciences has revived Edgeworth's project of measuring "true" utility. Quite tellingly, some behavioral economists are urging economists to recover the Benthamian foundations of the marginalist revolution (Kahneman, Wakker, and Sarin 1997).

This provides the background for a current debate that has been launched by Gul and Pesendorfer's (2010) critique of neuro- and behavioral economics. Gul and Pesendorfer strongly argue against the conflation of economics and psychology that underlies the research program of behavioral economics. Notably, they argue that economics is concerned with a set of issues quite distinct from those psychology is dealing with. In tackling these issues, economics uses quite different abstract structures (i.e. theories, models and concepts) than psychology as well as different kinds of evidence. In particular, they claim that the relevant informational basis for economics is legitimately restricted to choice data. Gul and Pesendorfer's general strategy is to shield mainstream economics from the critiques coming from behavioral economics regarding its assumptions about human behavior by claiming that these critiques are irrelevant since they target aspects which are *not* relevant from an economic perspective. A striking feature of Gul and Pesendorfer's manifesto is that it relies on a revealed preference approach endowed with the very properties rejected by Sen: first, they contend that the choice theoretic structure that underlies economics as a whole has purely behavioral foundations.[4] Second, they explicitly claim that individuals' choice behavior, i.e. their revealed preferences, is the relevant welfare criterion.[5]

[2] "[T]he faith in the axioms of revealed preference arises, therefore, not from empirical verification, but from the intuitive reasonableness of these axioms interpreted precisely in terms of preferences… the whole framework of revealed preference analysis of behavior is steeped with implicit ideas about preference and psychology" (Sen 1973, 243).

[3] It is worth noting however that it would probably be excessive to claim that all 20th neoclassical economists endorsed a revealed preference account of microeconomic concepts. For instance, the partial equilibrium analysis in the tradition of the Chicago School and especially the work of Gary Becker are not easily interpretable through a revealed preference stance.

[4] "In the standard approach, the term utility maximization and choice are synonymous. A utility function is always an ordinal index that describes how the individual ranks various outcomes and how he behaves (chooses) given his constraints (available options). The relevant data are revealed preference data; that is, consumption choices given the individual's constraints. These data are used to calibrate the model (i.e., to identify the particular parameters) and the resulting calibrated models are used to predict future choices and perhaps equilibrium variables such as prices. Hence, standard (positive) theory identifies choice parameters from past behavior and relates these parameters to future behavior and equilibrium variables" (Gul and Pesendorfer 2010, ??).

[5] "Hence, welfare is *defined* to be synonymous with choice behavior" (Gul and Pesendorfer 2010, ??, emphasis in original).

Gul and Pesendorfer's characterization of standard economics exemplifies what Hands (2013) labels *contemporary revealed preference theory* (CRPT). Hands (2013, 1087) suggests that CRPT is constituted by the three following claims:

a) CRPT uses consistency axioms like WARP to infer behavioral patterns on the basis of finite sets of choice data.
b) CRPT defines preferences solely in terms of choices and argues against the claim that preferences causally explain choices. Preferences are merely restatement of choice behavior.
c) CRPT is methodologically imperialistic, i.e. it claims to offer a general framework for choice theory in economics.

Claim (a) indicates that CRPT uses the revealed preference apparatus in an empirical rather than a theoretical perspective. While Samuelson's goal was to establish the formal equivalence between ordinal utility theory and revealed preference theory, contemporary revealed preference theorists rather follow the research programmed opened by Afriat (1967) who showed how to infer a set of utility functions from a finite choice data set. Claim (b) refers to the so-called "causal utility fallacy" (Binmore 2009): in a revealed preference perspective, Ann chooses *x* rather than *y not because* Ann prefers *x* to *y* (or because Ann's utility for *x* is higher than for *y*); rather, Ann prefers *x* to *y* because she chooses *x* rather than *y*. Another way to put this claim is that revealed preference theory is not expected to *causally explain* choice behavior. Claim (c) is the most significant one. It is a central one in Gul and Pesendorfer's critique of behavioral economics: basically, it states that doing choice-theoretical economics is to use revealed preference in conformity with the two previous claims. Any approach that does not follow these guidelines is not regarded as wrong per se; rather, it is simply regarding as belonging outside the domain of economics.

As I suggest above on the basis of Gul and Pesendorfer's text, CRPT builds on features of revealed preference theory that have been regarded as problematic from the start, including by well-known standard economists. Most of the time, the critiques' targets include the more or less rampant behaviorism that is thought to underlie the revealed preference approach. This is precisely here that NSE becomes relevant as it aims at providing CRPT with more solid methodological and philosophical foundations than pure and naïve behaviorism. Ross (2014a) is responsible for the label "neo-Samuelsonian" and also for the elaboration of most of its content.[6] He makes the following twofold claim (Ross 2011): i) Gul and Pesendrofer are right in arguing that the economist's concept of *choice* has nothing to do neither with how psychologists define the concept nor with its folk-meaning; however, ii) they are unnecessarily dogmatic in their uncompromising rejection of any evidential basis other than choice data and their claim that cognitive latent functions and processes do not matter in explaining choice behavior and market dynamics. As a consequence, NSE agrees with Gul and Pesendorfer that economics and psychology are two strictly separated fields. However, it does not follow them in their strict (and somewhat outdated) behaviorism. Indeed, Ross (2014a, 201) characterizes NSE as "anti-behaviorist anti-behavioralist". The former refers to the rejection of Gul and Pesendorfer's uncompromising behaviorism, the latter to the rejection of behavioral

[6] See also Clarke (2016), Dowding (2002) and Satz and Ferejohn (1994) for similar interpretations and defenses of revealed preference and rational choice theories.

economists' commitment to the redefinition of foundational economic concepts (choices, preferences) in a psychological or mentalist perspective.

3. Neo-Samuelsonian Economics and Economic Agency

The distinguishing characteristic of NSE thus conceived is located in its sophisticated and (somehow) counterintuitive account of economic agency. In a nutshell, NSE recognizes economic agency as an *interface pattern* between minds and environmental factors, especially social institutions such as markets. Economic agency is directly observable through choices, defined as *behavioral patterns that are directly responsive to opportunity costs through changes in incentives and the information distribution across a population*. Though choices are observable at an individual level, economics is mostly concerned with population-level choices, i.e. choices that "are identified only statistically, as tendencies observed over runs of instances, usually in aggregated sets of agent responses observed across a population, when incentives change exogenously or are manipulated by an experimenter" (Ross 2011, 222). Choices, and thus preferences, are inferred through the use of the theoretical apparatus of revealed preference theory (for choices under certainty) and expected utility theory (for choices under risk and uncertainty).[7] Basically, assuming that choices are consistent with a set of axioms, it becomes possible to infer future and unobserved choices from finite data sets on the basis of which utility functions are parametrized. Alternatively, structural econometric techniques allow to infer from a finite set of choice data which is the best theoretical model of decision-making to account for the choices of a given population (or subset of).

The key point of this account of economic agency is the separation between the economic agent and the individual person. Indeed, economic agency only requires observed and incentive-sensitive choices on the one hand, and consistency with choice-theoretic axioms on the other hand. There is no presumption that any choice should necessarily be related to a well-identified individual person. Behavioral patterns can also be ascribed to a whole population of persons or, at the opposite of the scale, to some of the "sub-personal selves" of a single person. Sup-personal selves and populations are thus also legitimate economic agents, at least as long as their choice behavior can be rationalized by some choice-theoretic account. This disentanglement between economic agency and personhood is of course unproblematic at the purely formal level. Choice functions and binary preference relations that are the subjects of the various axioms of choice theory do not need to be about *persons*' choices and preferences. Of course, this has been long ago recognized by economists who routinely ascribe utility functions to households and production functions to firms. However, the neo-Samuelsonian decoupling of economic agency from personhood goes farther. The stronger claim is that persons are not *prototypical* economic agents:

> "on the most parsimonious conceptual regimentation consistent with contemporary cognitive and behavioral science an individual cognitively and emotionally competent adult human only approximates agency in an indirect sense… although there are organisms that are indeed

[7] As I discuss in section 5 below, proponents of NSE do not regard expected utility theory as necessarily being the best theoretical account of choices under risk and uncertainty. Rank-dependent utility theory is also taken to be an at least as good account of experimental data. Quite the contrary, cumulative prospect theory (behavioral economists' pet theory of choices under risk and uncertainty) is generally argued to be an *inferior* account.

> paradigmatic agents – insects, for example – humans are not instances of such organisms" (Ross 2005, 241).

This claim is especially grounded on experimental results that strongly indicate humans' predisposition to make intertemporally inconsistent choices. Suppose that an economist is interested in the choices made by Ann across a more or less extended time period. Because Ann may have change her mind about some set of issue relevant for her choices and/or because she does not discount future reward exponentially, it is highly probable that Ann's choices will reveal inconsistent preferences. Obviously, this violation of consistency axioms makes difficult if not impossible to use Ann's previous (observed) choices to infer her future and/or unobserved choices. The behavioral economist's conclusion from this observation would be that Ann is non-rational and that her behavior should be accounted for by different, psychologically-grounded, decision-theoretic accounts. The Neo-Samuelsonian conclusion would rather be that Ann is not an authentic economic agent. Her behavior should rather be accounted for through a game-theoretic model of interactions of a community of agent corresponding to Ann's temporally located sub-personal selves.[8] In the model, each sub-personal selves is assumed to behave in accordance with choice-theoretic axioms, which in particular implies that their behaviors are amenable to a representation in terms of the maximization of a utility function.

The neo-Samuelsonian account of economic agency is anchored in two related philosophical doctrines (Ross 2014a, 236-51): ecological rationality and externalism about the semantics of mental states. The former refers to a view inherited from the Scottish enlightenment and especially put forward recently in the work of the experimental economist Vernon Smith (2009). It emphasizes that rationality is not necessarily a property possessed by individual beings but that may nevertheless emerge at the collective level. In economics, this view has been especially articulated by Hayek (1945) in his account of the market as an efficient information processor. The basic idea is that even though market participants may be boundedly rational and endowed with imperfect information, an appropriate institutional setup may result at the market level in a more or less close approximation to the standard assumptions about aggregate demand and efficiency results. Works such as Becker (1962) and Gode and Sunder (1993) provide great illustrations of efficient markets with a well-behaved demand function, in spite of the fact that participants behave in an almost random way. Ecological rationality is thus the product of a complex set of interactions between minds (e.g. market participants' desires, beliefs, preferences…) and social institutions sending signals (e.g. prices) such that the collective behavioral pattern meets efficiency criteria.

Externalism about the semantics of mental states (or semantic externalism for short) completes the neo-Samuelsonian account of economic agency. Semantic externalism has been notoriously defended on the basis of thought experiments by Putnam (1975) and Burge (1986) among others. It claims that the semantic content of a person's mental states (i.e. the truth-value of propositions about someone's desires, beliefs, wants…) is not fully determined by this person's intrinsic properties (the view of internalism). Rather, this content is necessarily related to the environment in which the person is embedded. In other words, the truth-value of propositions such as "Ann desires *x*" or "Bob believes that *y*" does not only depend on Ann's and Bob's intrinsic or internal states. The meaning of their attitudes is co-determined by the intrinsic states that realize in their brains *and* by the social environment with which they are interacting: "Just

[8] Ainslie (2010) and Schelling (1984) have famously informally suggested to model intertemporal inconsistency this way. Benabou and Tirole (2000) is a formal version of this idea.

as economic agency denotes an interface pattern between minds and markets, so the mind itself, as understood by externalists, is an interface pattern that describes systematic relationships between brains and socially related people" (Ross 2014a, 242). The conjunction of ecological rationality and semantic externalism thus indicate that relevant economic phenomena, especially market dynamics, are the result of the interactions between latent cognitive processes taking place in individuals' brains and market institutions.

Ross (2005) gives to this conjunction a particular flavor by augmenting semantic externalism with Daniel Dennett's (1989) account of intentionality and agency that Ross labels "intentional-stance functionalism". According to the latter, intentional states and their meaning are uncovered by taking the intentional stance. It consists in explaining and predicting the behavior of an entity (a human being, but also an animal or a machine) but ascribing to it intentional states such as desires or beliefs. The intentional stance has an apparently strong instrumentalist flavor that corresponds to what can be called the "Dennettian method" for behavioral explanation (Ross 2002, 154) or "methodological intentional-stance functionalism" (Ross 2005).

However, a purely instrumentalist reading of Dennett's account would indicate that belief attribution is merely based on a falsifiable theory of the mind and that we should give it up provided we are able to show that this theory is false or unnecessary (with respect to some parsimony criterion). This is precisely the eliminativists' position that Dennett rejects. Consider indeed hypothetical Martians who are observing Humans and are trying to predict our future on the basis of superhuman abilities making them equivalent to Laplacean super-physicists: "Our imagined Martians might be able to predict the future of the human race by Laplacean methods, but if they did not also see us as intentional systems, they would be missing something perfectly objective: the *patterns* in human behavior that are describable from the intentional stance, and only from that stance, and that support particular generalizations and predictions" (Dennett 1989, 25, emphasis in original). Dennett's point is that the intentional stance is not merely instrumental; it is the *only* way to observe *real* behavioral patterns. According to Dennett's "ontological intentional-stance functionalism" (Ross 2005), there is thus nothing more in the fact that the entity E has the belief that φ than the fact that E's behavior can be interpreted and predicted (by E itself or others) from the intentional stance through the ascription to E of the belief that φ. This is a form of realism, though a mild one since in many cases one's mental states will be partially indeterminate from the intentional stance (Dennett 1991).[9] Interestingly, as Dennett (1989, 58) notes himself, his "intentional system theory" has close connections with "already existing disciplines as decision theory and game theory, which are similarly abstract, normative, and couched in intentional language". In other words, choice-theoretic tools on which economists routinely rely are also ways to take the intentional stance and to ascribe intentional states to economic agents.

4. NSE and the Reconciliation Problem between Positive and Normative Economics

This section argues that the neo-Samuelsonian decoupling between economic agency and personhood leads to what McQuillin and Sugden (2012) have called a "reconciliation problem"

[9] This indetermination has a strong formal similarity with Quine's radical translation problem, as Dennett notes at several places.

between positive and normative economics. McQuillin and Sugden's point is different however as it concerns the implications of the normative turn of behavioral economics. Still, the ultimate consequences of these two distinctive reconciliation problems are highly similar: the standard preference-satisfaction welfare criterion that is at the core of mainstream welfare economics seems no longer appropriate. The main reason is that the preference-satisfaction criterion has been essentially grounded since the days of Pareto on a principle of consumer sovereignty. This principle provides the justification to take persons as the appropriate loci of the welfare analysis. However, it appears to be much less relevant once it is recognized that persons are not prototypical economic agents. At least, I shall argue for this in the section.

Though this is not directly to the main subject of this paper, it is worth briefly explaining why behavioral economics is confronted with a reconciliation problem. As I indicate in the introduction, the consumer sovereignty principle has traditionally been defended on the basis of two postulates: (i) individuals are rational (i.e. have consistent preferences according to a set of axioms) and (ii) individuals' preferences are constitutive of their well-being. Together, these two postulates provide a way to bridge positive and normative economics in a welfarist perspective: on the one hand, it makes possible to make welfare assessments of states of affairs on the basis of individuals' observed and hypothetical (inferred on the basis of choice axioms) choices (from the positive to the normative); on the other hand, assuming that individuals are rational, it is possible to infer from welfare assessments what the individuals' choices would be given some constraints. It is clear that postulate (ii) is purely axiological: it is an account of the nature of well-being grounded on a more general ethical theory. Postulate (i) however is at least partially empirical: it is a statement about the world that, given the underlying rationality norms, may be true or false.[10] Moreover, there is a sense in which postulate (i) provides some justification to postulate (ii), and this justification is the core of the consumer sovereignty principle as it is tacitly understood by economists. The fact that preference-satisfaction is constitutive of welfare is especially plausible *because* individuals are rational, i.e. individuals have preferences with characteristics that are sufficient (and maybe necessary) for these preferences to be constitutive of welfare.

To see this, just consider some of the classical axioms that are imposed to the strict binary preference relations in ordinal utility theory (the same reasoning is easily applied to the axioms of revealed preference theory). A first axiom is that strict preferences should be asymmetric, i.e. if x is strictly preferred to y, then y cannot be strictly preferred to x. A second axiom is that strict are negatively transitive, i.e. for any z, if x is strictly preferred to y, then either x is strictly preferred to z or z is strictly preferred to x, or both. Together, these two axioms guarantee that the related weak preference relation is complete and transitive (Kreps 1990, 23). As it is well-known, these axioms are necessary and sufficient for the preference relation to define a complete ordering of the elements belonging to any set X of states x. Adding a continuity assumption, the preference relation can be represented by a set of utility functions unique up to any positive transformation. Therefore, these axioms guarantee that the preference relation is equivalent to some (ordinal) quantitative measure. Considering the normative domain now, the welfare economist wants to be able to order states of affairs belonging to a set X according to a

[10] I am not saying that there are no moral facts and that axiological or ethical statements can never be true or false. My point is totally agnostic regarding the status of moral realism and moral cognitivism. What matters is that the first postulate is unequivocally partially empirical, i.e. subject to being characterized as true or false through empirical inquiry.

"(strictly) better than" relation where "x is strictly better than y for agent i" if and only if i's welfare is higher in x than in y. To make such comparative evaluations, it is necessary to assume that the "better than" relation defines an ordering and even maybe that welfare can be (at least) ordinally measured. Hence, the "better than" relation should satisfy the same axioms than the preference relation. Under this assumption, we therefore know that it is possible to obtain a "welfare ordering" from one's preference ordering over X; we only have to define welfare as preference satisfaction by substituting the "better than" relation for the preference relation.

Positive behavioral economics strongly indicates that real persons in experimental settings consistently violate axioms of utility theory. If we take these results for granted, it is easy to see why there is a reconciliation problem in light of the preceding paragraph. Basically, individuals' preferences *do not form a complete ordering* and, assuming that the "better than" relation must define a welfare ordering, it is no longer possible to assume that the preference relation and the "better than" relation are isomorphic. Behavioral economists interested in normative matters have proceeded in three different ways to maintain a bridge between positive and normative economics. The most radical solution is the "back to Bentham" approach of Kahneman and others (Kahneman, Wakker, and Sarin 1997) who distinguish between "decision utility" and "experience utility". The former corresponds to the utility notion that one finds in ordinal and expected utility theory, i.e. utility is nothing but a mathematical device to represent choices and preferences. The latter is a measure of various forms of hedonic states, especially instant happiness. Crucially, for these behavioral economists, only the latter is normatively relevant. Obviously, this implies that the preference-satisfaction view of welfare is simply given up. A second possibility is to claim that welfare only corresponds to the satisfaction of "rational" or "laundered" preferences. This is for instance the approach explicitly endorsed in Sunstein and Thaler's (2003) defense of libertarian paternalism. A related but formally different approach is to view an individual's choices as the aggregation of the choices made by her sub-personal selves and to discount the choices of the least rational selves in the social evaluation. This is for instance how Bernheim and Rangel (2009) proceed in their "behavioral welfare economics".

I am not interested here in evaluating the plausibility of these different approaches. I want however to emphasize that the reconciliation problem faced by behavioral economics has only two solutions: either to simply give up the standard view of welfare as preference-satisfaction, or to reconsider the status of the individual person as the locus of welfare analysis. Though the reconciliation problem confronted by the NSE is not quite the same than for behavioral economics, it has similar solutions, or so it seems. Given what has been said in the preceding section, the reconciliation problem in NSE does not come so much from the rejection of (i) than from the fact that (ii) is implausible.[11] Obviously, this is due to the fact that preferences are not intrinsic attributes of human persons within NSE: on the one hand, preferences are consistent choice patterns that are ascribed to economic agents rather than to persons; on the other hand, because of semantic externalism, preferences cannot be purely subjective and intrinsic properties of economic agents. The content of preferences is partially determined by the agent's environment and cannot be directly ascribed to persons. Clearly, once we acknowledge this point, it seems difficult to hold that preferences are constitutive of *individuals*' welfare.

[11] Neo-Samuelsonian economists are generally skeptical regarding the external validity of experimental results coming from behavioral economics. Notably, it is emphasized that if individuals have the possibility to engage in some learning process, most of the behavioral biases quickly evaporate. See Levine (2009).

If preferences and their content are not intrinsic properties of persons, the consumer sovereignty principle seems far less appealing. However, that does not imply that the preference-satisfaction account of welfare is automatically disqualified. Rather, as for behavioral welfare economics, we may hold that only *some* preferences are axiologically relevant. Or we may find and provide an alternative foundation for the preference-satisfaction account: for instance, one may argue that preferences revealed by choice patterns are good proxies of individuals' welfare, even though individuals are not authentic economic agents. The point remains that because the consumer sovereignty principle is no longer available, a disconnection between economic agency and normative agency is introduced. Before exploring possible solutions to this disconnection, I illustrate briefly how it manifests in experimental studies taking place within NSE.

5. Behavioral Econometrics and Welfare Assessments in Neo-Samuelsonian Experiments

A great illustration of the disconnection between economic and normative agency in NSE is given by a set of experimental studies aiming at estimating the consistency of risk and time preferences in a pool of subjects.[12] These studies are specifically designed to capture and measure preferences heterogeneity within the population and use the so-called technique of "mixture models" estimation to make welfare assessments of individuals' choices. This section describes the salient features of this experimental approach and argues that it is ambivalent regarding the role of agency in the articulation between the estimation of utility functions and the resulting welfare assessments.

The technique of model mixture estimation can be put in direct relationship with the neo-Samuelsonian assumption that choices and preferences are interface patterns resulting from the interaction between institutions and latent cognitive structures. Consider for instance a utility function representing an agent's risk preferences over a set of lotteries. In a neo-Samuelsonian perspective, the estimation of this utility function (i.e. the determination of the value of its free parameters) should be made against the agent's choices over pair of lotteries belonging to this set. As an illustration, we may assume the following standard CRRA (constant relative risk aversion) utility function:

(1) $$u(x) = \frac{x^{1-r}}{1-r}$$

Here, $r \neq 1$ is a parameter determining the shape of the utility function and x any positive number corresponding to monetary gains. Therefore, $r > 0$ entails concavity of the utility function and (under expected utility maximization) risk aversion. Assuming that the agent maximizes her expected utility, the value of r can be estimated against choice data over pairs of lotteries to determine her risk attitude. The point is that the estimated utility function corresponds to *latent* risk preferences, i.e. dispositions to choose among a set of lotteries on the basis of which it is possible to predict with more or less accuracy the agent's choices. A utility function is thus nothing but a choice pattern that can be tested against actual and observed choices. It does not indicate *how* the agent is actually choosing (what is happening in her mind and brain) but the pattern results from the interaction of the agent's cognitive processes and the various

[12] Some of the most significant studies are (Andersen et al. 2008), (Andersen et al. 2014), (Harrison and Ng 2016), (Harrison and Ross 2016). For a general account of the methodology of "behavioral econometrics", see Harrison (2017).

institutional features that determine the agents' incentives and opportunity costs (i.e. monetary outcomes, objective probabilities).

The assumption that an agent is a CRRA expected utility maximizer is however somewhat arbitrary. Not only an agent's choices may correspond to another class of utility functions (for instance utility functions entailing constant *absolute* risk aversion), but it is also quite possible that she is not an expected utility maximizer. For instance, she may weight objective probabilities of states of nature depending on the rank of the corresponding outcomes and/or weight monetary outcomes depending on whether there are framed as losses or gains. Moreover, most of the time, risk preferences are estimated for a population of agents against pooled choice data. In this case, we should allow for a possible heterogeneity of utility functions in the population. That is, while some agents may behave as CRRA expected utility maximizers, other agents may exhibit other kinds of risk attitudes better reflected by a utility model corresponding, for instance, to rank dependent utility theory (Quiggin 1982) or cumulative prospect theory (Tversky and Kahneman 1992). Mixture models estimation then uses maximum likelihood elicitation procedures to determine the distribution of choice models in the population. This permits to proceed to "horse races" between competing models to estimate which one best fit any subset of choice data. The most interesting aspect of this approach with respect to the neo-Samuelsonian account of economic agency is that it "treat choices, rather than sets of choices by specified individuals, as the data against which models are estimated… thus the traditional a priori assumption that a person is a utility function for economic modeling purposes need not be maintained" (Ross 2014a, 155-6). In particular, the heterogeneity of utility functions within the population may be imputed either (or both) to the fact that not all individuals instantiate the same utility functions through their choices *or* to the fact that the same individuals may instantiate different utility functions in different circumstances. In the latter case, personhood and economic agency are disentangled.

Harrison and Ng's (2016) experimental study of the expected welfare effects of insurance provides a great illustration of this general methodology as well as of its implication for welfare assessments. The authors proceed through two steps. In the first step, elicitation of the subjects' risk attitudes and determination of the distribution of utility functions within the population are done by asking each subject to make 80 binary choices between lotteries with objective probabilities. It is determined whether each subject's choices are better described by an expected utility or a rank dependent utility model and, if the latter, which kind of probability-weighting function offers the best fit. The second step consists in assessing the welfare effects of subjects' choices over insurances on the basis of the risk preferences elicited in the first step. Subjects are asked to express their binary willingness to pay for insurance against a potential loss. Each subject is offered a range of premium amounts to insure her against a 0.1 probability of a determinate loss occurring. For each premium and each subject, it is possible to calculate the expected welfare gain from purchasing insurance. Expected welfare gains are simply measured in term of ex ante consumer surplus, i.e. the difference between the certainty equivalents of the (expected or rank-dependent) utility function with and without insurance. Obviously, it is rational to buy an insurance if and only if the consumer surplus is non-negative. Harrison and Ng's study indicates that subjects suffer from significant losses as a result of paying for insurance when they should not (according to their elicited risk preferences) and not buying insurance when they should (again according to their elicited risk preferences). A similar result occurs in a study of Harrison and Ross (2016), this time regarding investment decisions in financial products.[13]

Beyond these specific results, experimental studies like Harrison and Ng's have several methodological and philosophical implications for the relationship between positive and normative economics in NSE. First, it should be noted that the welfare assessments that are made are nonsensical unless it is assumed that estimated risk preferences elicited through binary choices over simple lotteries carry over choices made over more complex insurance or financial products. Otherwise, the use of certainty equivalents to compute consumer surplus would be impossible. Implicitly, it thus assumed that there is a relative structural stability across choice environments, i.e. the underlying cognitive processes lying behind choice patterns are such that model mixtures estimations made in a given choice environment remain valid in a different choice environment.

This assumption has philosophical significance given the fact that, even though in Harrison and Ng's study (and also in Harrison and Ng's one), utility functions are explicitly ascribed to *individuals*, this is neither a methodological nor a conceptual necessity. Indeed, a second implication in a normative perspective is that risk preferences are not *subjects'* risk preferences in the usual sense. The structural stability referred above is not the stability of the subjects' brain processes but rather of the relationship between *heterogeneous choice patterns* and different incentive structures. This has noteworthy consequences for normative debate over paternalism in economics as it blurs the recent distinction made between "nudges" and "boosts": the former refers to the action of a "choice architect" trying to steer individuals toward choosing options that are deemed best for them by manipulating their environment; the latter quite the contrary directly act on individuals by educating them such as to make them able to make the best choices. However, semantic externalism makes difficult to entertain this distinction most of the time individuals are not aware of "their" preferences, as the latter are the simultaneous product of cognitive processes, incentive structures and also interpretations by experimenters (Harrison and Ross 2016).

The last and most important implication concerns the relative ambivalence of welfare assessments in studies like Harrison and Ng's one. The very use of the well-known normative concept of consumer surplus obviously indicates that the individual person remains a natural locus for the welfare analysis. However, from a normative perspective neo-Samuelsonians are more interested in the *aggregate* consumer surplus rather individual consumer surplus. Moreover, consumer surplus can be technically computed without necessarily ascribing utility functions to individuals but rather to choice patterns referring to sub-personal selves or groups of individuals. But then, the point made in the preceding section is more relevant than ever: if preferences are not individuals' preferences, on which basis can they be considered as welfare-relevant? In the preceding section, I argue that preference-satisfaction is a plausible account of welfare thanks to the consumer sovereignty principle but that this principle is not available for

[13] Subjects in Harrison and Ross (2016) were divided into two groups. In the control group, subjects were not given any investment advice while in the treatment group, subjects follow an "education program" helping them to choose in which products to invest. Welfare effects were measured by calculating, for each subject, the difference between the certainty equivalent of the optimal portfolio conditional on risk preferences and the certainty equivalent of the actual portfolio chosen. It appears that subjects in the treatment group suffer from significantly *lower* welfare losses than those in the control group.

neo-Samuelsonians. The next two sections explore two possibilities to decipher this problem and to set the ground for an authentic neo-Samuelsonian welfare economics.

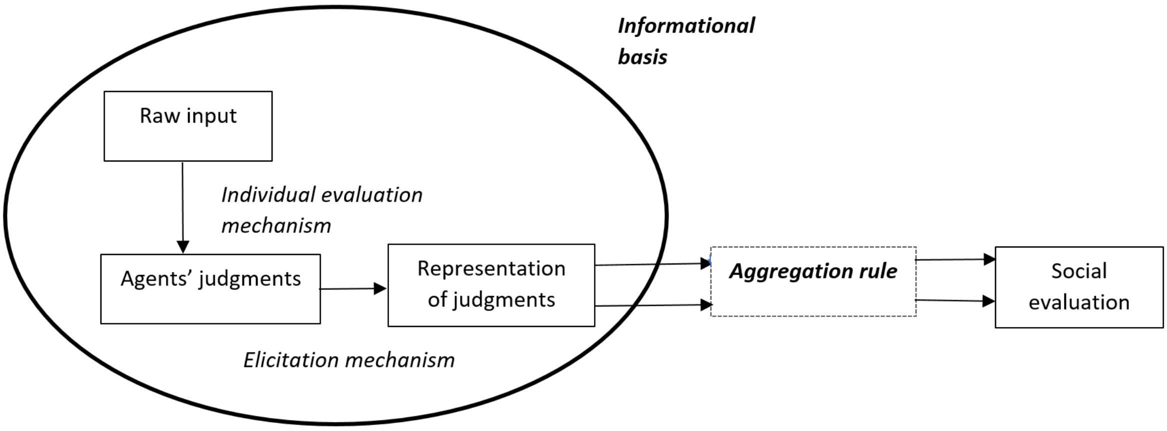
6. Neo-Samuelsonian Welfare Economics and Formal Welfarism

This section discusses a first possible reconciliation between positive and normative economics in a neo-Samuelsonian perspective. This reconciliation is only partial however because it actually consists taking normative economics as being *part* of positive economics. This view transpires in Gul and Pesendorfer's account of the relationship between revealed preference theory and welfare economics:

> "Economists used the revealed preference of individuals as a welfare criterion because it is the only criterion that can be integrated with positive economic analysis… Economists use welfare analysis to identify the interest of economic agents and to ask whether existing policies can be interpreted as an expression of those interests or whether the understanding of the institutional constraints on policies remains incomplete. This use of welfare analysis requires the standard definition of welfare" (Gul and Pesendorfer 2010, 25).

Two points are noteworthy. First, as I suggest in the second section, Gul and Pesendorfer's characterization of the relation between positive economic analysis and welfare analysis builds on the very conflation between choices, preferences and personal welfare that Sen (1973) argued against. Second, Gul and Pesendorfer explicitly claim that welfare economics is a tool for positive analysis and that the preference-satisfaction account of welfare is justified by this objective. This is of course significant since it means that the preference-satisfaction view needs not be grounded on the consumer sovereignty principle. The basic argument is the following: if it is acknowledged that the purpose of welfare analysis is to allow for a *positive* account of the institutional constraints on policies (i.e. whether and why policies are efficient), then it is almost necessary to define welfare as preference-satisfaction because preferences (defined as choice patterns) are the only legitimate informational basis in positive economics. Any other account of welfare will not do simply because it requires a kind of information that is not part of the economist's informational basis (e.g. subjective measure of happiness or neural firing rates obtained by neuroscientific methods). Though this claim may seem somewhat counterintuitive at first sight, it makes perfect sense in the perspective of the so-called "formal welfarism" approach in social choice theory (Fleurbaey 2003). Sen (1979, 468) famously characterizes welfarism as the doctrine according to which "the judgment of the relative goodness of alternative states of affairs must be based exclusively on, and taken as an increasing function of, the respective collections of individual utilities in these states". Welfarism is thus "essentially an informational constraint for moral judgments about states of affairs" (Sen 1979, 472) since it indicates that only individual utilities are welfare-relevant. Sen's original characterization of welfarism implicitly assumes that utilities entering as input in the social welfare functional (a) are ascribed to individual persons and (b) that they correspond to a measure of some subjective mental states. The latter point is of course consistent with Sen's critique of revealed preference theory and his argument that preferences have to be understood in a psychological or mentalistic sense (see section 2 above). Formal welfarism retains the definition of welfarism as an informational constraint on moral judgments but relax both (a) and (b). Indeed, Sen's welfarism correspond to what Fleurbaey (2003) characterizes as "real welfarism", the latter being a subset of formal welfarism (Gharbi and Meinard 2015).

Fig. 1: The social choice model of normative analysis



We can achieve a better understanding of formal welfarism if we put it in relation with what I will call the "social choice model of normative analysis" (see fig. 1).[14] The social choice model essentially builds on the distinction between two key concepts: the *informational basis* and the *aggregation rule*. The informational basis corresponds to the set of information on the basis of which states of affairs are evaluated. The aggregation rule specifies how the information collected should be used to generate the social evaluation. In particular, it specifies restrictions on the resulting social evaluation given the content of the information basis, e.g. if everyone prefers some state of affairs *x* to another state *y*, then *y* cannot be evaluated as socially better than *x*.[15]

The informational basis and the aggregation rule are the two basic ingredients of the social choice model of normative analysis but additional features are generally needed. In particular, *within* the informational basis, we may need to distinguish between what we can call the 'raw input' and its representation. The former corresponds to the brute information as it is directly available to the normative economist. It may consist for instance of data on the distribution of wealth or health states vectors defined along standard measures. However, this raw input is generally not what is aggregated. Rather, aggregation proceeds on the basis of the *individual evaluations* of the states of affairs characterized by the raw input as well as on their representations through some formal devices (e.g. preference orderings, utility functions). Consider the following simple example: say we want to evaluate a set of states of affairs according to the amount and the distribution of wealth within them. We collect actual and/or hypothetical data on wealth distribution. This is what corresponds to the raw input of the informational basis. However, what is to be aggregated are not these data but rather the individuals' evaluative judgments on states of affairs as characterized by wealth distribution. To do so, assume that individuals are able to form some kind of judgment on the possible states of affairs and that these judgments are somehow 'well-behaved'. For instance, a basic requirement for a judgment to be well-behaved would be that if individual *i* judges state of affairs *x* better than state of affairs *y*, and *y* better than state *z*, then *i* should judge *x* better than

[14] My characterization of the social choice model is inspired but sensibly differed from Hausman (2010).

[15] The distinction between the informational basis and the aggregation rule is quite standard in welfare economics and social choice theory. Bear in minds however that it is mostly a heuristic. In practice, it is not always possible to disentangle both. This is particularly obvious when one looks at the details of the various axiomatization in social choice theory. Several commonly used axioms seem to determine *both* the informational basis and the aggregation rule at the same time. The various Paretian and independence axioms are a case to this point.

z. Second, individuals' judgments should be elicited through some procedure. This elicitation procedure should be sufficient to measure individuals' judgments in a relevant way and to permit their representations through a formal device. Ultimately, these measurements that will be aggregated to form the social evaluation. The informational basis is thus obtained through the raw input and the combination of at least two mechanisms: (i) an individual evaluation mechanism through which we figure out that individuals are able to form their judgments over states of affairs on the basis of the raw input; (ii) an elicitation mechanism through which we assume we are able to represent the individuals' judgments in the required form.

Conceived as an "informational constraint", formal welfarism is essentially concerned with the nature and the content of the informational basis. Both are constrained by what is known in the social choice literature as the "strong neutrality" axiom which in substance states that nonwelfare information (i.e. not represented in the profile of utility functions) is irrelevant in the ordering of social alternatives (Blackorby, Bossert, and Donaldson 2005). The strong neutrality axiom is itself obtained by combining three more basic axioms: unlimited domain (all possible profiles of utility functions can be aggregated), Pareto indifference (if two profiles of utility functions are identical, then the two corresponding social alternatives are ranked together) and binary independence of irrelevant alternatives (the ranking of two social alternatives depends only on the – welfare and nonwelfare – information associated with those two alternatives only). Within these constraints, formal welfarism is totally permissive regarding i) what is to be aggregated and ii) the nature of agency on the basis of which aggregation proceeds. Both are obviously directly relevant for the determination of the informational basis.

Consider the former point first. As it is standard in social choice theory, a social welfare functional can be defined as a mapping F: $\mathcal{U} \rightarrow \mathcal{O}$ of a vector of n individual utility functions $u = (u_1, \ldots, u_n)$ defined over some relevant domain $\mathcal{U}$ onto a social ordering $o \in \mathcal{O}$ of a set of social states X. The social ordering is thus a function of the possible profiles of individual utility functions figuring in the relevant domain. While the function F is defined over profiles of individual utility functions and thus aggregates *utilities*, the precise nature of the latter is left undefined. That is, individual utility functions may represent almost anything that one deems as being axiologically relevant. To be more specific, the individual utility functions may represent the degree of preference-satisfaction as it is standardly assumed in welfare economics. But they may also measure what is sometimes called "experienced utility" in cases where one's axiology is grounded on a mental state account of well-being. Individual utility functions may also be grounded on various "objective list" theories of well-being, including those endorsing a notion of capabilities. Finally, they may also represent the agents' (subjective) judgments regarding well-being. Regarding the second point, nothing in the formalism requires that the subscripts in the vector of individual functions refer to persons *qua* agents. Indeed, as suggested by Mirrlees (1982), the individual utility functions may be ascribed to sub-personal selves.

The latter possibility is of course especially interesting in the neo-Samuelsonian perspective where economic agency and personhood are disconnected. In particular, it opens the way for a philosophical treatment that follows Derek Parfit's (1984) suggestion that personal identity and even mere psychological connectedness between sub-personal selves are axiologically irrelevant. In this extreme case, one may agree with Parfit that while the scope of morality is widened (morality extends to the relationship between the selves of the same person), its weight is reduced: since distributive concerns between sub-personal selves of the same person are generally disregarded, the irrelevance of personhood may be argued to justify to disregard them also for the relationship between persons. As Parfit suggests, this reductionist account of personhood provides a new defense for utilitarianism, the latter consisting in adding the utilities of sub-personal selves rather than of persons:

> "the Utilitarian View may be supported by, not the conflation of persons, but their partial disintegration. It may rest upon the view that a person's life is less deeply integrated than most of us assume. Utilitarians may be treating benefits and burdens, not as if they all came within the same life, but as if it made no moral difference where they came. And this belief may be partly supported by the view that the unity of each life, and hence the difference between lives, is in its nature less deep" (Parfit, 1984, 335-6).

I do not intend to claim that the account of economic agency that is constitutive of NSE necessarily leads to Parfit's "complete" utilitarianism. Indeed, this depends on what is assumed regarding the cardinality and the comparability of utilities and the ordinalist revolution that is at the roots of NSE is well-known for being responsible to the high skepticism economists have had about both.[16] My point is elsewhere: formal welfarism and the underlying social choice model of normative analysis provide a way to bridge positive and normative economics in a way that is both compatible with NSE and Gul and Pesendorfer's view of the role of welfare analysis. Of course, this implies that normative economics is *part of* the positive analysis in the sense that both the locus of the welfare analysis (the economic agents) and the underlying account of welfare as preference-satisfaction are determined by the nature of the positive – neo-Samuelsonian – analysis. In particular, in the neo-Samuelsonian interpretation of the social choice model, agents' judgments are actually mere consistent choice patterns that are not necessarily ascribed to individual persons. This view seems to be perfectly consistent with the kind of welfare analysis illustrated in section 5.

7. Neo-Samuelsonian Welfare Economics, Personhood and Value-Based Life Satisfaction

The characterization of NSWE in the preceding section builds on the assumption that normative agency is determined by economic agency, i.e. what is welfare-relevant depends on what and who an economic agent is. However, this clashes with most persons' well-entrenched moral views that there is something morally special about *persons*. This moral intuition is not necessarily affected by the neo-Samuelsonian contention that persons are not prototypical economic agents. Indeed, the same can be said regarding the reconciliation problem that occurs in behavioral economics: in spite of the fact that normative behavioral economists strongly argue that not all individuals' preferences should be taken into account in the welfare analysis, autonomy of the individuals is nonetheless regarded to be normatively significant, as witnessed by the debates over so-called "libertarian paternalism". To use a distinction introduced by Sen (1991), the fact that individuals are not prototypical economic agents does not imply a *welfare-perspective* applied to economic agents in normative economics, but quite the contrary is

[16] This is actually a quite complicated issue that I cannot discuss in this paper. Both von Neumann-Morgenstern's and Savage's accounts of expected utility theory indicate that it is possible to generate *cardinal representations* of preferences in a behaviorist perspective. Similarly, interpersonal utility comparisons can be elicited in certain circumstances on the basis of observed choices only. However, cardinal representation of preferences and cardinal preferences should not be confused, as noted by Weymark (2005) among others.

compatible with an *agency-perspective* applied to persons. The purpose of this section is to offer an argument for this claim in the context of NSE.

Ross (2014a, 303-12) is aware of the tension resulting from the separation of economic agency with normative agency. One of the implication of NSE and its account of economic agency is that "descriptive individualism" is false; but somewhat counterintuitively, "normative individualism is explained by the fact that descriptive individualism is false" (Ross 2014a, 311, emphasis removed). Ross's claim is grounded on a theory of narrative construction of personhood extensively developed in his 2005 book (Ross 2005). The starting point of this theoretical account is again the neo-Samuelsonian claim that persons are not prototypical economic agents. A person's behavior exhibits multiple different utility functions across choice situations and time frames. These utility functions correspond to different and partially conflicting interests, different risk attitudes and time preferences, and so on. The result is almost necessarily some degree of choice inconsistency across time and across choice situations. Narrative construction processes function such as to reduce as far as possible this inconsistency by making the person's behavior and rationalizable and minimally predictable by others but also by the person herself. The role of a narrative is precisely to give to every member of a given community a plausible rationale for the person's choices and attitudes.

More specifically, Ross argues that each person should take the intentional stance toward others and toward herself by trying to build a rationale for everyone's behavior. This process takes place in an institutional environment made of social institutions consisting in particular in signaling devices. Ross (2005, 291-316) formalizes the process as a nested (conceptual but not necessarily temporal) sequence of three types of games. *G''*-type games through which "biological utility functions" are determined are played between authentic economic agents. *G'*-type games take place between players that are already human persons (each with their biological utility function) and who are strangers for each other. In these games, persons send and receive signals about a set of characteristics (each person's preferences, normative expectations and commitments, …) that will contribute to determine the kind of *G*-type games that they will be ultimately playing. The latter are played by fully self-narrated persons whose ability to coordinate on an efficient equilibrium will depend on the stability and the consistency of expectations that have emerged from *G'*-type games. Institutions provide coordinating devices that most of the time will allow the players to reach a "satisficing" solution. This implies in particular that each person is able to stabilize herself over some intrapersonal bargaining equilibrium taking the form of a credible and consistent self-narrative. Thus, in Ross's account, economic agency and personhood are ultimately reconciled.

In this perspective, market institutions are unique in their incentivizing properties for behavioral consistency. Beyond sheer luck, economic successes and failures are largely the result of the persons' (and also organizations') ability to behave in the most consistent way. Inconsistent individuals are in particular vulnerable to money pump and Dutch book mechanisms. While this is true in all institutional contexts, this is especially the case on the marketplace where public signals (i.e. prices and opportunity costs) facilitate the exploitation of behavioral inconsistencies. Financial markets provide the best theoretical and empirical illustration of this point: funds and individuals making inconsistent decisions due to for instance erroneous probabilistic risk assessments will be exploited by arbitrageurs and quickly driven out of the market.[17] This is precisely such kind of error-exploitation mechanisms that is responsible for the ecological rationality that markets tend to exhibit. The point is that market institutions incentivize individuals to behave *as if* they were authentic economic agents with well-behaved utility functions, while in other kinds of social situations these incentives are significantly weaker. This account of personhood is fully consistent with semantic externalism: persons are the product of complex interactions between latent cognitive processes (i.e. "minds") and social institutions sending signals and generating incentives. One is a person only through her own and others' eyes in a specific set of social contexts.

Nevertheless, the pressure for behavioral consistency that characterizes our modern economies should not be overestimated. Even on the marketplace, persons and organizations are not always authentic economic agents. This is not only due to cognitive biases, sheer lack of awareness or social choice problems, but also because modern economies are inherently guided by a Schumpeterian dynamic of creative destruction which by its very nature is very hard to predict. Of course, innovations cannot be anticipated before they actually arise and it follows that it is impossible for instance for today's consumers to already know what their (revealed) preferences will be in the next few years.[18] More generally, no one is able to anticipate what our societies will look like in two or three decades and thus it cannot be expected (neither descriptively nor normatively) that individuals will/must behave in a fully consistent and predictable way. The market discipline cannot change this fact, if only because market mechanisms *are* the cause of many social and technological disruptions. Acknowledging that persons cannot therefore be authentic economic agents in spite of narrative construction processes, it remains to determine why they should nonetheless be considered as the loci of welfare analysis (and more generally of normative economics). I suggest that a plausible answer can be found in recent philosophical accounts of well-being that emphasize the role of values and authentic happiness regarding the way persons conduct their life. Crucially, these accounts are fully compatible with the above view of personhood in terms of narrative construction.

Value-based life satisfaction accounts of well-being have been developed over the last three decades to overtake the difficulties related to traditional subjective theories of well-being, especially preference-satisfaction and mental state ones (Tiberius and Plakias 2010). They build on the postulate that well-being is constituted by the fact of living a life that one takes as being "good" or "satisfactory" relatively to one's own values about what matter. Obviously, these accounts assume and indeed require that persons are capable of self-scrutiny and reflexivity over their values and are able to put both their choices and exogenous events in relations with these values. In other words, it is assumed that persons are reasonable beings who, even though they are prone to mistakes, errors and inconsistencies, can reflect over their choices and what happen to them. Most of the time, happiness is taken to be constitutive of well-being. However, happiness is here conceived neither in hedonic terms nor as a pure mental state. The non-identification of happiness with hedonic mental states is notably the key feature of Graham

[17] As suggested by Ross (2014, 234-5), this may provide an explanation to Ainslie's (2010) observation about the surprising fact that monetary interest rates in modern economies are far below the level they should have if people were actually hyperbolically discounting future rewards. Since hyperbolic discounting leads to intertemporal inconsistencies, the market discipline strongly incentivizes investors and savers to avoid behaving on its basis.

[18] Harari's (2017) popular "prospective history" of 21st century humanity singles out the emergence of "Dataism" as the new world religion. Dataism downplays the normative importance of persons and instead venerates algorithms able to predict people's wants and behaviors even *before* individuals are themselves aware of them. In the context of this paper, Dataism could be characterized as the ideology viewing persons as authentic economic agents, or perhaps more disturbingly seeking to transform persons in authentic economic agents.

Sumner's (1996) "welfarist" account of well-being. Sumner submits that well-being is constituted by *authentic happiness*:

> "The theory I shall defend does not simply identify well-being with happiness; additionally, it requires that a subject's endorsement of the conditions of her life, or her experience of them as satisfying or fulfilling, be authentic. The conditions for authenticity, in turn, are twofold: information and autonomy. Welfare thus consists in authentic happiness" (Sumner 1996, 139).

The information condition plausibly requires that the person's happiness is not grounded on false beliefs, ignorance, misperception or self-delusion. Autonomy can be regarded as the requirement that the values and reasons on the basis of which you are more or less satisfied with your life should be your own. This does not mean that these values and reasons are not shared across the population or that you should not have acquired them by interacting with other persons. Rather, it demands that they are not due to mechanisms such as social conditioning or indoctrination that deprive you from your self-assessing reflexive abilities. In other words, *these* values and reasons are your own because you would be willing and able to defend and endorse them in a publicly deliberative context and under rational scrutiny. Of course, there is no clear-cut method to determine whether or not one's values are authentic in this sense. Socialization is such that values and reasons are never completely our own. The point is that autonomy requires that the values that are guiding your self-assessment of your life should indeed be your own and being distinguishable from those of other persons or of groups. Moreover, you should be willing to rationally endorse the values on the basis of which you self-narrate your life. Otherwise, your narrative could be regarded as fictitious and unreliable by others, and your self-assessment of your life welfare-irrelevant.

Thus, persons' (revealed and perhaps inconsistent) preferences are important in the welfare analysis not because of a consumer sovereignty principle. They matter because they *depend on* and *realize* underlying values that determine what one takes to be a "good", "satisfactory" and "fulfilling" life. While preferences-as-consistent-behavioral-patterns can be ascribed to economic agents that do not identify with individual persons, the same is less plausible for values. The key differences is that while one may have preferences she is not aware of and/or unable to intelligibly articulate, values are at the core of narrative construction processes. In other words, we *justify* to others and to oneself our behavior on the basis of our values. By nature, values thus hold on an intertemporal scale but also across choice situations. A person whose values do not satisfy these requirements will be regarded by others as erratic at best, insane at worst. It may be worth insisting at this stage that my understanding of the "value" concept is fully externalist: values are not more "in people's mind" than preferences (or any other intentional states) are. Like preferences, values are identified through behavioral manifestations and result from the interactions of minds and social institutions. The difference however is that in NSE, preferences correspond to choice patterns that correlate with opportunity costs. That means that preferences are specifically singled out using the tools and concepts of choice-theoretic economics. Values are not responsive to change in opportunity costs in the same way. As a result, they are more difficult to identify on a purely behaviorist basis. They nonetheless transpire through persons' behavior, including (and maybe especially) through *verbal* behavior. If this point is granted, it has an interesting implication regarding the relevant informational basis for welfare economics in a neo-Samuelsonian perspective: Gul and Pesendorfer's claim that only choice data are legitimate within economics is not only arbitrary but also mistaken: it is simply impossible to make welfare assessments without asking *why* persons' make the choices they make. A neo-Samuelsonian *welfare* economics should be far more tolerant and open-minded regarding the extent of the informational basis.

8. Conclusion

This paper has discussed the implications of the neo-Samuelsonian account of economic agency for welfare analysis and more generally normative economics. The distinctive property of NSE is its disconnection between economic and normative agency. As a result, I have argued that it faces a reconciliation problem between positive and normative economics similar to the one that occurs in behavioral economics. I have explored two possible routes for reconciliation. The first, directly in line with Gul and Pesendorfer's (2010) suggestion that welfare analysis depends on positive economic analysis, is the approach corresponding to the social choice model of normative analysis and culminating in the notion of "formal welfarism". Such a route is compatible with extreme form of normative views completely downplaying the axiological importance of persons, such as Parfit's complete utilitarianism. The second possibility builds on Ross's account of personhood as narrative-construction processes. To lead to a full normative account, I have argued that it could be completed by recent theories of well-being emphasizing the role of values in happiness and life satisfaction. This route sensibly diverges from a pure behaviorist understanding of welfare economics, as it indicates that the economists' informational basis should be widened to include values and reasons underlying people's behavior.

References

Afriat, S. N. 1967. "The Construction of Utility Functions from Expenditure Data." *International Economic Review* 8 (1): 67–77.

Ainslie, George. 2010. *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*. Reissue edition. Cambridge: Cambridge University Press.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2008. "Eliciting Risk and Time Preferences." *Econometrica* 76 (3): 583–618.

———. 2014. "Discounting Behavior: A Reconsideration." *European Economic Review* 71 (October): 15–33.

Becker, Gary S. 1962. "Irrational Behavior and Economic Theory." *Journal of Political Economy* 70 (1): 1–13.

Benabou, Roland, and Jean Tirole. 2000. "Self-Confidence and Social Interactions." Working Paper 7585. National Bureau of Economic Research. http://www.nber.org/papers/w7585.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *The Quarterly Journal of Economics* 124 (1): 51–104.

Binmore, K. G. 2009. *Rational Decisions*. Princeton University Press.

Blackorby, Charles, Walter Bossert, and David J. Donaldson. 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge University Press.

Burge, Tyler. 1986. "Individualism and Psychology." *The Philosophical Review* 95 (1): 3–45.

Caplin, Andrew, and Andrew Schotter. 2010. *The Foundations of Positive and Normative Economics: A Handbook*. Oxford University Press.

Clarke, Christopher. 2016. "Preferences and Positivist Methodology in Economics." *Philosophy of Science* 83 (2): 192–212.

Colander, David. 2007. "Retrospectives: Edgeworth's Hedonimeter and the Quest to Measure Utility." *The Journal of Economic Perspectives* 21 (2): 215–26.

Dennett, Daniel C. 1991. "Real Patterns." *The Journal of Philosophy* 88 (1): 27–51.

Dennett, Daniel Clement. 1989. *The Intentional Stance*. MIT Press.

Dowding, Keith. 2002. "Revealed Preference and External Reference." *Rationality and Society* 14 (3): 259–84.

Fleurbaey, Marc. 2003. "On the Informational Basis of Social Choice." *Social Choice and Welfare* 21 (2): 347–84.

Gharbi, Jean-Sébastien, and Yves Meinard. 2015. "On the Meaning of Non-Welfarism in Kolm's ELIE Model of Income Redistribution." *Journal of Economic Methodology* 22 (3): 335–53.

Gode, Dhananjay K., and Shyam Sunder. 1993. "Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality." *Journal of Political Economy* 101 (1): 119–37.

Gul, Faruk B. and Wolfgang Pesendorfer. 2010. "The Case for Mindless Economics." In A. Caplin & A. Schotter (eds.), *The Foundations of Positive and Normative Economics*, Oxford University Press, 3-39.

Hands, D. Wade. 2013. "Foundations of Contemporary Revealed Preference Theory." *Erkenntnis* 78 (5): 1081–1108.

Harari, Yuval Noah. 2017. *Homo Deus: A Brief History of Tomorrow*. HarperCollins.

Harrison, Glenn. 2017. "The Methodologies of Behavioral Econometrics." In M. Nagatsu and A. Ruzzene (eds.), *Philosophy and Interdisciplinary Social Science: A Dialogue*, London: Bloomsbury, forthcoming.

Harrison, Glenn W., and Jia Min Ng. 2016. "Evaluating the Expected Welfare Gain from Insurance." *Journal of Risk and Insurance* 83 (1): 91–120.

Harrison, Glenn and Don Ross. 2016. "Varieties of Paternalism and the Heterogeneity of Utility Structures." *CEAR Working paper*.

Hausman, Daniel M. 2010. "Valuing Health: A New Proposal." *Health Economics* 19 (3): 280–96.

———. 2011. *Preferences, Value, Choice, and Welfare*. Cambridge University Press.

Hayek, F. A. 1945. “The Use of Knowledge in Society.” *The American Economic Review* 35 (4): 519–30.

Hédoin, Cyril. 2016. “Sen’s Criticism of Revealed Preference Theory and Its ‘Neo-Samuelsonian Critique’: A Methodological and Theoretical Assessment.” *Journal of Economic Methodology* 23 (4): 349–73.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. “Back to Bentham? Explorations of Experienced Utility.” *The Quarterly Journal of Economics* 112 (2): 375–406.

Kreps, David M. 1990. *A Course in Microeconomic Theory*. Princeton University Press.

Levine, David. 2009. *Is Behavioral Economics Doomed? The Ordinary versus the Extraordinary*. Cambridge: Open Book Publishers.

McQuillin, Ben, and Robert Sugden. 2012. “Reconciling Normative and Behavioural Economics: The Problems to Be Solved.” *Social Choice and Welfare* 38 (4): 553–67.

Mirrlees, James A. 1982. “The economic uses of utilitarianism”. In Amartya Kumar Sen & Bernard Arthur Owen Williams (eds.), *Utilitarianism and Beyond*, Cambridge University Press. 77-81.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.

Plakias, Alexandra and Valerie Tiberius. 2010. “Well-Being.” In J. Doris and the Moral Psychology Research Group (eds.), *The Moral Psychology Handbook*, Oxford: Oxford University Press, 401-431.

Putnam, Hilary. 1975. “Language and reality.” *Mind, language and reality* 2: 272-290.

Quiggin, John. 1982. “A Theory of Anticipated Utility.” *Journal of Economic Behavior & Organization* 3 (4): 323–43.

Ross, Don. 2002. “Dennettian Behavioural Explanations and the Roles of the Social Sciences.” In A. Brooks and D. Ross (eds.), *Daniel Dennett*, Cambridge University Press, 140-83.

———. 2005. *Economic Theory And Cognitive Science: Microexplanation*. MIT Press.

———. 2011. “Estranged Parents and a Schizophrenic Child: Choice in Economics, Psychology and Neuroeconomics.” *Journal of Economic Methodology* 18 (3): 217–31.

———. 2014a. *Philosophy of Economics*. Palgrave Macmillan.

———. 2014b. “Psychological versus Economic Models of Bounded Rationality.” *Journal of Economic Methodology* 21 (4): 411–27.

Samuelson, P. A. 1938. “A Note on the Pure Theory of Consumer’s Behaviour.” *Economica* 5 (17): 61–71.

Satz, Debra, and John Ferejohn. 1994. “Rational Choice and Social Theory.” *The Journal of Philosophy* 91 (2): 71–87.

Schelling, Thomas C. 1984. “Self-Command in Practice, in Policy, and in a Theory of Rational Choice.” *The American Economic Review* 74 (2): 1–11.

Sen A. K. 1977. “Rational Fools: A Critique of the Behavioral Foundations of Economic Theory.” *Philosophy and Public Affairs* 6 (4): 317–44.

Sen, Amartya. 1973. “Behaviour and the Concept of Preference.” *Economica* 40 (159): 241–59.

———. 1979. “Utilitarianism and Welfarism.” *The Journal of Philosophy* 76 (9): 463–89.

———. 1991. *On Ethics and Economics*. Reprint edition. Oxford, UK; New York, NY, USA: Wiley-Blackwell.

Smith, Professor Vernon L. 2009. *Rationality in Economics: Constructivist and Ecological Forms*. 1st ed. Cambridge University Press.

Sunstein, Cass, and Richard Thaler. 2003. “Libertarian Paternalism Is Not An Oxymoron.” SSRN Scholarly Paper ID 405940. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=405940.

Tversky, Amos, and Daniel Kahneman. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty* 5 (4): 297–323.

Weymark, John A. 2005. “Measurement Theory and the Foundations of Utilitarianism.” *Social Choice and Welfare* 25 (2–3): 527–55.