A new definition of and role for preferences in positive economics

Bart Engelen*

Tilburg Center for Logic, Ethics, and Philosophy of Science (TiLPS), Tilburg School of Humanities, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, The Netherlands

(Received 1 September 2016; accepted 17 January 2017)

Positive economic models aim to provide truthful explanations of significant (aspects of) economic phenomena. While the notion of ‘preferences’ figures prominently in micro-economic models, it suffers from a remarkable lack of conceptual clarity and rigor. After distinguishing narrow homo economicus models (self-interest maximization) from broader ones (preference satisfaction) and rehearsing the criticisms both have met, I go into the most promising attempt to date at addressing them, developed by Hausman. However, his definition of preferences as ‘total comparative evaluations’, I argue, plays into the general disregard that economists have for human psychology. My alternative definition of preferences as ‘overall comparative evaluations’ – and hence as one of the many factors that influence people’s behavior – allows for more adequate causal explanations of people’s dutiful, committed, and norm-guided actions. Against Hausman but in agreement with Sen, it also allows for (motivated) counterpreferential choice.

Keywords: homo economicus; preference; total comparative evaluation; counterpreferential choice; Hausman; Sen

Introduction

Preferences are at the core of economics. Economists claim it to be part of their success in truthfully predicting and explaining (aspects of) economic phenomena (positive economics) and analyzing the desirability of specific policies and institutions (normative economics). Here, I will focus specifically on positive economics, which I take to develop models, and theories that generate substantive knowledge about complex real-world phenomena. This can be done by deriving falsifiable predictions that can be tested empirically (Lipsey, 2008). While one can, in my view rightly, argue that a lot of economists only pay lip service to this idea, making only vague and casual remarks (if any) about the relation between their models and real-world phenomena (Sugden, 2000, p. 6), the increasing prominence of experimental and behavioral economics reveals this idea to be alive, kicking, and actually gaining momentum. Relying on empirical evidence, ‘the behavioral approach aims at developing more realistic psychological models to replace rational homo economicus’ (Guala, 2008). Preferences play a pivotal role in these and many other economic models.

Despite its importance, the notion ‘preferences’ suffers from a remarkable lack of conceptual rigor. In what follows, I critically analyze the work of Daniel Hausman, who provides by far the most extensive and promising attempt at defining preferences.

*Email: b.engelen@uvt.nl

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
and clarifying how they can (and cannot) function in positive economics. First, I contrast ‘preference satisfaction’ with the narrower assumption of ‘self-interest maximization’. Next, I go into Hausman’s definition of preferences. He not only clarifies the use of preferences in conventional economic practice but also justifies it. While I agree he accurately captures how preference-based models in fact model individuals, I stress important disadvantages of this strategy. In response, I then provide an alternative definition of preferences and argue for a new way of understanding the relationship between preferences on the one hand and motivations such as duties and commitments on the other hand. I argue why my approach – which stresses that such motivations can influence choices ‘directly’ rather than ‘indirectly’ via preferences – can deal with the problems that Hausman’s raises. In the final section, I formulate a conclusion.

Self-interest maximization and preference satisfaction

Let me first distinguish between two views in philosophy of economics on how economic models can truthfully explain and represent people’s motivations and actions.

(1) According to isolationism, of which the main proponent is Mäki (1992, 2009), economic models can and should be viewed as analogous to experiments: they abstract from minor causes in order to isolate major causes and their effects. If this isolation succeeds, models provide partial but significant truths about (essential aspects of) human behavior and interaction.

(2) According to fictionalism, of which the main proponent is Sugden (2000), economic models can and should be viewed as analogous to novels: they provide counterfactual but credible worlds. If these model worlds are significantly similar to the real world, economists can make inductive inferences from the former to the latter.

Both views stress that a similarity between (specific features of) model worlds and the real world is possible and needed for explanatory purposes (Mäki, 2009, p. 30). Now, how have (simple and abstract) economic models and theories tried to capture the major causes of (complex and concrete) actions by real-life people in their attempts to causally explain those actions (see also Reiss, 2012)?

An important strand of economic thought models agents as self-interest maximizers. Mill (1836/1874, section V 38) famously defined political economy as the discipline that views man ‘solely as a being who desires to possess wealth, and who is capable of judging of the comparative efficacy of means for obtaining that end’. While economists know that real-world behavior is caused by a wide variety of motivations, they abstract from this, ‘treating the main and acknowledged end as if it were the sole end’ (Mill, 1836/1874, section V 38). Their simplified stories do not provide the whole truth, the argument goes, but can still truthfully explain people’s actions by ‘idealizing’ or ‘isolating’ the most significant causes.

Self-interested homo economicus is, however, a highly unrealistic representation of actual people. Growing empirical evidence reveals the widespread existence of anomalies which suggest that self-interest maximization is often not ‘the main and acknowledged end’. Such evidence comes from both experiments (e.g. altruistic punishment: why punish another if this only brings about costs and no benefits to oneself?) and everyday life (e.g. voting behavior: why go out to vote if the impact of a single vote is negligibly small?). The motivations that cause people to deviate from self-interest maximization in
Ultimatum Games, on election days and in many other instances – reciprocity, duty, internalized norms, et cetera – are not properly understood as minor or collateral. In addition, they are often not some kind of self-interest in disguise. While it is always possible to claim that any given action serves the agent’s self-interest (Sen, 1977, p. 322), such a post hoc rationalization and ad hoc construction of self-interest often does not refer to what actually causes the agent’s action.

A lot of economists have broadened the homo economicus model by dropping the ‘self-interest maximization’ assumption altogether and replacing it with ‘utility maximization’ or ‘preference satisfaction’. In response to both empirical and theoretical criticism – by authors such as Herbert Simon, Ronald Coase, Daniel Kahneman and Amartya Sen (1977) – conventional practice is to construct complete and transitive preference functions that give rise to consistent, utility-maximizing behavior. The idea is that ‘utility maximization should be a central tool in analyzing human behavior, even if humans are not self-regarding’ (Gintis, 2006, p. 7; see also Schmidtz 1993). In fact, this strategy dates back to 19th century economists such as Léon Walras (1874/2003, p. 65) and Alfred Marshall (1890: book I, chapter II, §3) who already warned against the ‘mistaken’ belief that economics has ‘no concern with any motive except the selfish desire for wealth’.

Dropping the self-interest assumption, recent theories model players as being ‘inequality averse’, preferring both more money and more equality (e.g. Fehr & Schmidt, 1999) or as having ‘prosocial preferences’, such as altruistic preferences (players care about payoffs to others), utilitarian preferences (players care about total payoff), Rawlsian preferences (players care about the minimum payoff) or a combination of these (e.g. Charness & Rabin, 2002). Other models postulate preferences for reciprocity, fairness or altruistic punishment (Rabin, 1993; Fehr & Gächter, 2002).

**Hausman’s definition of preferences**

Let us take a closer look at this notion of ‘preferences’. In the most systematic and comprehensive treatment of the different understandings, definitions and usages of preferences, Hausman (2005, pp. 33–35; 2012, pp. 1–3) convincingly argues that preferences are not to be defined in terms of (1) self-interest or expected advantage, (2) desires, likings or enjoyment comparisons or (3) actual or hypothetical choices. Instead, preferences – as most economists use the term – are best defined as total comparative evaluations and hence as rankings of alternative choice options in terms of all considerations that the person finds relevant (Hausman, 2005, pp. 37–38; 2012, pp. 3–4).

This broad definition is in line with mainstream economic practice, Hausman (2012, p. 36) argues, for example in analytical game theory, where payoffs often no longer represent benefits or dollars gained but ‘utilities’ or ‘utils’ which are shorthand for preference rankings. If an alternative has a higher payoff, it means that the agent ranks it higher than and thus prefers it over its alternatives.

According to the standard ‘BPC model’ in decision theory, an agent’s actions are to be explained in terms of his beliefs, preferences, and constraints (Gintis, 2009, p. 1; Hausman, 2012, pp. 36–37). With beliefs covering the agent’s cognitive elements, preferences capture the agent’s affective and motivational elements. Instead of distinguishing different motivations such as egoism, altruism, and a sense of duty, these are all lumped together as factors that influence preference rankings: ‘this panoply of factors that influence people’s actions – their plans and passions, their whims and wishes, their tastes and temptations – must all find their place within an agent’s preferences and..."
beliefs’ (Hausman, 2012, p. 57). Of the factors that Hausman (2005, p. 37; 2012, p. 134) takes to influence people’s preferences are: beliefs, tastes, longings, emotions, (dis)likes, ideals, commitments, cognitive heuristics and biases, akrasia, habits, social norms, moral principles, other people’s wishes, arguments, and so on. I will argue later on that quite a few of these motivational factors are best understood not as part of but as rival to people’s preferences.

The main advantage of constructing complete and transitive preference functions is that it allows economic models to become more tractable and mathematically amenable. Locating preferences close to action, economists can also employ the axiom of ‘choice determination’: ‘Among the alternatives they believe to be available, agents, will choose one that is at the top of their preference ranking’ (Hausman, 2012, p. 15). Assuming other interfering factors to be absent enables economists to build (game and decision theoretic) models that predict and explain action: ‘only the conception of preferences as total subjective comparative evaluations permits game theory and expected utility to serve their predictive and explanatory roles’ (Hausman, 2012, p. 65; see also Hausman, 2005, pp. 45–49). Preferences as total comparisons thus allow economists to hold on to ‘game theory and preference-based prediction and explanation’ (Hausman, 2012, p. 69) and actually make them more informative (Hausman, 2012, p. 73).

Before going into my criticism of Hausman’s definition of preferences, let me simply note here that this will be based on the idea – clearly shared by Hausman as well – that causal explanation is one of the (many) purposes that economic models can and should serve (see also Mäki, 2009; Reiss, 2012; Sugden, 2000). After all, my argument is that, when models along Hausman’s lines take preferences as total comparative evaluations, they will often get the causal story wrong and thus fail to causally explain individual actions. Of course, if the model is to serve some other purpose, which it surely can, Hausman’s definition may still provide for a good modeling strategy.

Criticizing Hausman’s definition of preferences

While I think Hausman’s definition of preferences as total comparative evaluations adequately captures mainstream economic practice, I thus think it reveals a problem. In my view, it facilitates post hoc and ad hoc stipulations of preference rankings and utility functions that simply rationalize the agent’s actions. When an agent is seen choosing X, economists can always argue that, given her beliefs and constraints, she simply preferred X. Instead of making risky hypotheses, such a strategy enables economists to ‘explain’ away whatever seems anomalous. People who are seen voting are simply said to have a preference for expressing their political opinion.

According to Croson and Gächter (2010, p. 128), ‘any single observed behavior can be consistent with the theory; after all, we simply define a utility function (…) for which the observed decision is the outcome of its maximization’. Indeed, Hodgson (2013, p. 73) remarks, ‘the utility function (…) is an ex post rationalization of behavior rather than an explanation of its causes.’ Rationalizing actions by means of all-encompassing notions like ‘preferences’ as all-things-considered evaluative comparisons creates a black box that obscures what is of real explanatory interest.

By starting from ‘given’ preferences all things considered it becomes practically impossible to account for human action in terms of real processes operating in the human psyche. Casting over human motivation the veil of the utility representation we abstract from too much that is relevant for motivating human behavior. (Güth & Kliemt, 2007, p. 145)
One should not be a full-fledged Popperian to see the problem here. If every behavior can be fitted into a preference ranking or utility function\(^1\), the latter is no longer doing any real explanatory work. It will be the auxiliary assumptions about what exactly agents prefer in specific circumstances that yield interesting implications (Blaug, 1992, p. 232). Hausman (2012, p. 44) acknowledges this point.

If economists want to say more about choice among some set of alternative actions than that people choose what they prefer, they need to say something substantial about what determines what people prefer. In treating preferences as total comparative evaluations, economists have essentially trivialized the connection between final preference and choice.

Whatever explanatory or predictive force preference-based models have, comes from the stipulation of substantive assumptions about the content of preferences, not from stipulations about their structure (such as completeness, transitivity and convexity) (Brennan, 2007, p. 111). Think of theories that stipulate that people prefer more equal outcomes (‘inequality aversion’) or more payoffs for others (‘prosocial preferences’). Or take Thomas Schelling’s classic ‘checkerboard model’ (Schelling, 1969). The dynamics of this spatial segregation model are based on the players’ postulated mildly discriminatory preferences. Only by assuming that each player prefers a mixed neighborhood but also prefers more than one-third of his direct neighbors to be like himself, does Schelling’s model predict segregation.

The notion of preferences as all-things-considered comparisons lumps together all underlying motivational factors and thereby obfuscates what is of interest to whoever aims to explain human behavior. While this notion of preferences does not make it impossible for economists to ask what exactly motivates people, it enables them to shy away from this issue. The default strategy of practicing economists has indeed been to give up on causal explanation of individual action (Sugden, 2000, pp. 11–12).\(^2\) Largely in line with Milton Friedman’s dictum that each ‘theory is to be judged by its predictive power’ (Friedman, 1953, p. 8), most contemporary economists do not care about the extent to which their models truthfully represent major causal factors of human actions. They simply take people to behave ‘as if’ they were satisfying their highest-ranked preferences, relieving them from a deeper analysis of people’s actual motives. Insofar as causal explanation is one of the (many) purposes economic models can serve, I believe this is regrettable evolution that is furthered by an understanding of preferences as all-encompassing comparative evaluations.

Now, Hausman (2012, p. 65) considers it to be an advantage that, ‘by treating preferences as total rankings, economists can separate the use of the word “preference” from substantive views about what preferences depend on’. In contrast, I believe it facilitates a superficial re-description of behavior that neglects human psychology altogether. It allows economists to insulate themselves from disciplines that do provide more substantive views of people’s inner motivational lives. Whatever factor motivates a specific behavior, economists reason, it can be incorporated into a broad notion of preferences. In theory, because preference-based models no longer suffer from the narrow-minded tunnel vision of assuming self-interest maximization, they should be more open to insights from other behavioral sciences such as psychology, sociology, or biology (Gintis, 2009, p. 241). In practice, however, they allow economists to further isolate themselves from these disciplines and largely ignore the task of building empirically informed preference and utility functions. This evolution needs to be reversed if economic models are to causally explain individual action and thus ‘describe the causal mechanism responsible for it’ (Reiss, 2012, p. 43).
In my view, an alternative definition of preferences encourages economists to tackle head-on the roles different motivations play in different circumstances and to engage in interdisciplinary efforts to explain human behavior. This alternative definition contrasts my position with that of Hausman and Sen. While all three of us stress the importance of economists engaging in adequate causal explanation of human behavior, we disagree on how they should go about it. While Hausman argues that economists should focus on the factors that form preferences (as all-things-considered rankings), Sen believes economists can employ multiple conceptions of preferences. I will take a middle road, arguing that economists should go for a single conception of preferences (pace Sen) but define it more narrowly as overall and not as total comparisons (pace Hausman). On my account, some motivations help form people’s preferences but other motivations influence behavior in a more direct way (and are therefore, pace Hausman, not best investigated within preference formation theories).

To be clear, my quarrel with Hausman does not revolve around the (descriptive or clarificatory) issue how economists use the notion of preferences. I agree with Hausman that, in mainstream economic practice, preferences generally function as total comparative evaluations. I disagree with him, however, on the (normative) issue whether this strategy is desirable. While Hausman (2012, p. xii) states that ‘preferences in economics are and should be total subjective rankings’, I believe that, while they are, they probably should not be.³ Hausman’s analysis of the definition and use of preferences in economics is thus not merely descriptive in nature but serves as a justification (or an excuse) for existing modeling practices in economics. While Hausman (2012, p. 57) claims that incorporating ‘the panoply of factors that influence people’s actions (…) within an agent’s preferences and beliefs (…) is the way that economists understand choices and that it is a reasonable way to proceed’, I take issue not with the first (descriptive) claim but with the last (normative) claim. In what follows, I therefore analyze in more detail the downsides to the conventional strategy of understanding preferences and propose an alternative.

A new definition of preferences

The crucial issue is how to relate preferences to what we might call people’s motivations: the forces or influences that cause someone to do something.⁴ While Hausman (2005, p. 48) rightly agrees with Sen on ‘the importance of modeling the intricate thought processes individuals go through’, we disagree on how this should be done.

As shown before, Hausman understands motivations as factors that influence and form people’s preferences. If preferences are all-things-considered comparisons, then whatever motivates a person is simply incorporated in the formation of his preferences.⁵ According to Hausman (2012, p. 71), economic models ‘do not deny that human agency and motivation are complicated matters. They instead locate those complications in theories of preference formation and modification rather than in theories of choice.’ Economists should look at psychological and other empirical evidence to take on the hard task of finding ‘good theories of how people form and modify their preferences’ (Hausman, 2012, p. 131).

However, I believe this fails to understand how motivations like duties and commitments can cause people to act. To get a better grasp of how these motivations can influence behavior, I propose to redefine preferences as ‘overall’ rather than ‘total’ comparative evaluations. In this narrower conception of preferences, ‘people regard some of the factors that affect their evaluation of alternatives as competing with
preferences rather than as influencing preferences’ (Hausman, 2012, p. 3). In this view, preferences are partial rankings that include some of the motivational factors that cause human behavior but not all. Schematically, the difference between Hausman’s approach and mine is the following (ignoring the role of beliefs and constraints).

| Hausman's view:                                      | My view:                                    |
|-----------------------------------------------------|---------------------------------------------|
| all motivations (such as desires, norms, duties, ...) | some motivations (such as desires, …)       |
| total preference ranking                            | overall preference ranking                  |
| final choice                                        | final choice                                |
| judging, comparing, evaluating                      | judging, comparing, evaluating              |
| choice determination                                | other motivations (such as cravings, duties, norms, commitments, …) |

Before analyzing more fully how the causal impact of some motivations – such as cravings, duties, norms, and commitments – can bypass that of preferences, think of following examples.

- Doing what one believes to be the right thing even though one would rather not;
- Respecting another person’s priorities even though one would prefer that person giving in;
- Abiding by a collective decision even though one prefers a different outcome.

In each of these instances, Hausman can say that, if the agent in question decides to do X (e.g. do his duty), this simply means she has the all-things-considered preference to do X. Similarly, if the agent decides to do something else, Hausman can claim that her total comparative evaluation and thus her preference has changed. A sense of duty, like any other possible motivation, simply gets incorporated into the preference ranking. After all, ‘a total comparative evaluation takes into account every consideration the agent judges to be relevant’ (Hausman, 2012, p. 2).

In my view, this suffers from two problems. First, as mentioned before, the notion ‘preference’ is no longer doing any real work here. The predictive or explanatory force comes from the factors that form the preferences. Second, to the extent that it suggests a mental exercise of considering and ranking potential alternatives, it gives a misleading account of what is actually going on in people’s minds. I will focus on this second objection in what follows.

Now, if one understands preferences as overall and thus partial comparative evaluations – as I do – it is possible that an agent chooses an action without changing her preference. At the moment of choice, people can, for example, feel obliged or be seduced to act against her preference. Some motivational factors, such as a sense of duty, have a different structural relationship to and more direct influence on choices,
breaking the connection between preference and choice (and thus violating the assumption of ‘choice determination’).

Now, the definition of preferences as total comparative evaluations leaves hardly any conceptual space for counterpreferential choice. The only way in which people can act against their preferences is when they have false beliefs or make cognitive mistakes. Of course, I can choose something that goes against my preference for a glass of water when I mistake a glass of poison for water and decide to drink it. In contrast, my approach opens up room for ‘counterpreferential choice’ motivated not by cognitive but by affective factors that are not part of the agent’s preference ranking.

Two comments are relevant here. First, counterpreferential choice refers to conscious and deliberative choice, not to mere bodily movements and unconscious behavior. If I trip and break my leg, that goes against my preference but I clearly did not choose to. Second, the claim that people can consciously choose an option that (they know) goes against their own comparative evaluation, which I will defend here, is a stronger claim than the one Sen (1977, p. 327) already made in his seminal article ‘Rational Fools’, namely that one can act contrary to one’s preference in the sense of one’s expected advantage (see also: Hausman, 2005, pp. 39–42). The disconnect I am referring here is that between ‘comparative evaluation’ and ‘choice’ (which Sen stressed in his later writings), not that between ‘expected advantage’ or ‘welfare’ and ‘choice’ (which Sen stressed in 1977).

Relating preferences to cravings, duties, and commitments

In order to illustrate and justify my alternative definition, let us examine in detail three kinds of motivations that I believe can be modeled as bypassing rather than influencing preferences: (1) cravings, (2) duties, and (3) commitments.

First, the actions of most people who are weak-willed, addicted or swamped by sudden surges in craving cannot, in my view, be said to satisfy their preferences. While they have formed a comparative evaluation and judged what is best for them (and thus have constructed a preference ranking), these factors cause them to act against it. For reasons of space, I do not go into why I believe these examples still constitute instances of agency and choice rather than merely bodily reactions to biological factors. Instead, I will focus on two other motivations that are less controversial in this respect.

Second, think of duty, which has been defined as ‘the capacity to act against one’s preferences on the grounds that one ought so to act’ (Mansbridge, 1998, p. 155). Principled moral behavior, even when understood in terms of rationality and reasonableness, is not always properly understood in terms of preferences. To illustrate his non-Humean approach, Searle (2001, pp. 186–187) gives the example of someone ordering a beer, which in itself constitutes a reason for paying it, regardless whether or not that person prefers to pay for it. When visiting a bar, a set of norms is triggered – among which the norm that one should pay for one’s drinks. This norm and the accompanying sense of duty is not something people weigh off against other aspects relevant to the choice at hand (e.g. politeness versus money, taking into account reputation effects). A sense of duty is thus not always adequately modeled as a ‘consideration’ that enters into an all-things-considered comparative evaluation. It can bind one’s will and trump whatever consideration goes into such comparative judgments. While people often have preferences over what to have to drink (they may consider different options based on different criteria), they often do not have preferences over whether or not to pay for that drink. They pay for their drinks because they feel it is the right
thing to do, full stop. And if they were to actually form a preference, their sense of duty could motivate them to choose against it.\textsuperscript{7}

Third, a ‘duty’ can be understood as one form of what in the literature has become known as ‘commitment’. This dates back to Sen’s distinction between ‘sympathy’ (in which case one’s ‘concern for others directly affects one’s own welfare’; Sen, 1977, p. 326) and ‘commitment’ (in which case one’s concern for others motivates one’s action even if it would be against one’s own welfare). Here, commitment still fits an instrumental conception of rationality, with people choosing actions that best achieve their goals (which are no longer necessarily self-interested). In his later work Sen (1985, pp. 213–214; 2002, 35; 2005, pp. 6–7) thinks of committed action as non-instrumental in nature (see also: Schmid, 2005). Commitment, he argues, ruptures the link not only between an agent’s choices and her ‘self-centered welfare’ or ‘self-welfare goal’ (preferences as ‘expected advantage’) but also between her choices and her goals (preferences as ‘all-things-considered comparisons’). The idea is that people can (have good reasons to) take other people’s interests and goals as constraints on their own: “commitment can also involve violation of self-goal choice, since the departure may possibly arise from self-imposed restrictions on the pursuit of one’s own goals” (Sen, 1985, p. 348).

According to Searle (2001, p. 187), when I order a beer and drink it, I create a commitment that puts me under the obligation to pay for it. Commitments thus create reasons for action that are independent of the agent’s desires but that can still influence her choices (see also: Schmid, 2005, pp. 52–53).\textsuperscript{8} Let me reformulate Searle’s argument in terms of the explanatory role preferences play in my account. An adequate causal explanation of me paying my beer refers to the commitment I made when ordering and my sense of duty that derives from this. Deciding to pay is generally not the result of some all-things-considered comparative evaluation. Regardless whether I judge it best to pay for my beer or not, the commitment makes me feel it is not a genuine option not to. In this sense, duty and commitment influence choices in a way that is more akin to constraints than to preferences. Sen (2002, p. 43) actually talks about ‘self-imposed constraints’, distinguishing them from the externally originating “feasibility constraints”.

The presence of other people typically imposes constraints on what we perceive to be legitimate options. Sen (2007, p. 348) asks you to imagine occupying a window seat on a plane and being asked by your neighbor to pull down the shade so that he can play a computer game you find ‘plainly silly’. When you fulfil his request (even though you think he would be better off reading a book, Sen claims that you are not endorsing your neighbor’s goals or achieving whatever goal you might have. Instead, ‘you are just following a norm of good behavior (…), which is a self-imposed behavioral restraint you end up accepting in your choice of what to do’ (Sen, 2007, p. 349). Our recognition of other people and their goals does not imply that we identify with them or make them our own. Instead, we freely impose on our actions a ‘normative restraint’ (Sen, 2007, p. 353).

Schmidtz (1993, pp. 52–53) distinguishes here between ‘concern’ (I care about the interests of others, which enter into my preference function) and ‘respect’ (other people’s interests pose constraints on my preference function). The idea is the same: respecting others means that one sees there are limits on what one may legitimately do to them. Rather than a consideration to be weighed off against other considerations and influencing people’s choices via their total comparative evaluations, respect typically influences choices in a much more direct manner. Like commitment, respect is typically
not something that leads one to revise one’s preferences or comparative evaluations. Instead, they form self-imposed limits on the extent to which people (aim to) fulfill their preferences.

Broome (2001) uses a different terminology to express a similar thought. Preference rankings are close to what Broome calls the result of a process of weighing (pro tanto) ‘reasons’: considerations in favor of doing something that can but do not necessarily result in the decision to do it. In contrast, duty, commitment and respect are close to what Broome calls ‘normative requirements’, which cannot be outweighed by other considerations. Whereas reasons are ‘slack’, requirements are ‘strict’. If an agent does not live up to what, for example, her commitment to a moral principle requires from her, that constitutes a normative failure on her part. On my account, what causes the actions of people who do live up to what is required from them often cannot be adequately captured in terms of comparisons, rankings or all-encompassing considerations. They choose to do their duty, because they feel obliged or required to do so, full stop.

In his analysis of Sen’s views, Pettit (2005, p. 18) talks of ‘goal-displacing’ and not ‘goal-modifying’ commitment (see also: Cudd, 2014: 43), which actually fits my definition of (actions that go against) preferences as overall and thus partial comparative evaluations. Even though I prefer X (or have X as my goal), my commitment motivates me to do Y. When I am, for example, the trustee or the representative of another person or a group, my choice to carry out the trusteeship faithfully or serve other people’s interests is based on a binding commitment rather than on some revision of my preferences or goals that results from some all-things-considered evaluative judgment. What motivates my choice here are not so much my own preferences and goals (Sen, 2002, p. 216) but something that binds my will: a promise or pledge I made or my integrity as a friend or politician. As such, my preferences and goals are no longer the sole determinants of my choices (Sen, 2007, p. 350). Like Searle, Broome, and Sen (2005, p. 5) argue, there is nothing unreasonable or irrational about this at all.

Hausman’s reply

Hausman has two replies available to him. First, in line with Sen’s comments, he can take duties and commitments to imply that the agent considers certain options to be unavailable. Discussing Sen’s window seat example, Hausman (2012, p. 62) argues that ‘Sen’s acceptance of norms of courtesy rules out the possibility that he can leave his shade open’. Referring to the agent’s (beliefs about her) constraints enables Hausman to hold on to (1) preferences as all-things-considered rankings and (2) the ‘choice determination’ axiom that agent chooses her highest-ranked preference among the alternatives she believes to be available. If (Sen believes that) leaving open the shade is simply not an available option, ‘then, among the alternatives allowed by the constraint, he did what he most preferred’ (Hausman, 2013, p. 222).

I agree with Lehtinen (2013, p. 208) ‘that Hausman’s argument is not very persuasive here’. The reason why relates to the difficulty of specifying the options available to the agent (as she sees it) (Pattanaik 2013, pp. 217–218). People – such as Sen in the example – can realize fully well that the option is available but simply do not consider pursuing it. While I know that I can walk out of a bar without paying my beer and thus believe this option is available, my sense of duty discourages me from doing so.⁹

In fact, Hausman (2005, pp. 40–41) prefers a second strategy that models commitment as one of the factors that forms the agent’s preferences. If my sense of duty implies that I do not want to leave without paying my bill, this simply means it is a
determining factor in the formation of my all-things-considered preference. ‘In arguing that commitment, as counterpreferential choice, is precluded by a notion of preferences as total comparative evaluations, I argued that commitment should be understood as one of the factors determining total evaluation rather than as competing with preferences’ (Hausman, 2012, p. 71).

The underlying idea is to deny the claim by Sen (2002, p. 35) that self-goal choice can be violated (Pettit, 2005, pp. 20–21). According to Schmid (2005, p. 56), ‘one cannot pursue the other’s goals without making these goals one’s own’. Pettit and Schmid thus agree with Hausman that, while commitment can motivate an agent to act against her self-interest (preference narrowly defined as expected advantage), it cannot motivate her to act against her preferences (broadly defined as total comparative evaluations).

**Replying to Hausman’s reply**

Together with Sen and Schmidtz, however, I claim (1) that commitment can motivate people to act against their preference and (2) that this is in line with the common sense understanding of ‘preference’ and ‘action’. First, in Sen’s window seat example, one cannot reasonably claim that Sen aims to realize or comes to value his neighbor’s goal (of playing a silly computer game). Re-describing it in terms of some other goal (such as being a polite person) does not work either: Sen does not pull down the shades in order to uphold etiquette rules or to be some kind of person.

Second, pace Pettit (2005, p. 21), I believe ‘common-sense psychology’ is not congruent with but diverges from the view ‘that rational agents act out of a concern for maximizing the expected realization of their goals’. In my view – and Hausman (2012, pp. 3–4) acknowledges this – defining preferences as overall comparisons actually stays close to the everyday meaning of ‘preferences’, situating them further away from choice than Hausman’s definition does. While Hausman (2005, p. 45) believes his own definition ‘modestly extends’ everyday understandings of the word, he is aware that it ‘does not conform to ordinary usage of the word “preference.”’ (Hausman, 2005, p. 36). The difference between Hausman’s and the more everyday usage that I defend, is the extent to which preferences contrast with duties and commitments: not in Hausman’s terminology, a lot in everyday language. Whatever relevance ordinary language or folk psychology has in this respect, I believe it favors my approach over Hausman’s and Pettit’s. Hausman (2012, p. 17) himself seems to admit this much.

Now, to be clear, I am not claiming here that it is impossible to model self-imposed constraints or deontological obligations using ‘as if’ preference rankings. After Sen (2002, pp. 190–191) grants this possibility as well (see also: Cudd, 2014, p. 52), however, he adds that such modeling practices would only ‘mimic’ the agent’s actions instead of adequately representing the psychology that motivates these actions: ‘The as if preference works well enough formally, but the sociology of the phenomenon calls for something more’. In my view, Sen more accurately captures the phenomenology of what actually goes on in the abovementioned examples. This accuracy – of the motivational pluralism that Sen’s notion of commitment introduces – is important for whoever wants (economic models) to causally explain people’s behavior (Cudd, 2014, p. 55).

When Sen (2002, p. 304) criticizes economists who see ‘people as “rational fools”’, as undiscriminating and gross thinkers, who choose one all-purpose preference order’, his concern is thus that the latter does not adequately model actual people’s psychology. The criticism here is that the notion of ‘preference’ fails to capture specific aspects of, for example, moral motivation and action. People motivated by duties and
commitments either do not engage in the mental processes required to form a comparative evaluation or do engage in them but sometimes find that their resulting preferences are trumped.

While I agree with Sen’s idea that motivated counterpreferential choice is possible, I want to strengthen it by defining preferences in a single and coherent way. While Sen is lenient and accepts multiple conceptions of preferences, I think a strict definition of preferences as overall comparative evaluations specifies more clearly how duties and commitments can go against evaluative judgments and how they can influence behavior directly. In my view, this increases the explanatory power of preference-based models, since it opens up the black box which has enabled economists to lump together all possible motivations.

While complete and transitive preference functions assume that all aspects of a choice setting can be weighed off against each other and all alternatives can be ranked in terms of ‘betterness’, people can sometimes refuse to do this (and for good reason). Think of mother Sophie (in Sophie’s Choice) who has to choose whether her daughter or her son lives. These are options over which one simply cannot form a preference. Less dramatic but not less present in everyday life are the norms, rules, principles and values people embrace. When we uphold them, we generally act non-instrumentally and quasi-automatically, which is opposed to the instrumental and deliberate nature of forming and satisfying preferences.

In my view, Philip Pettit’s model of the virtual homo economicus captures the underlying psychology best (Pettit, 1995). What motivates people in most circumstances is not some conscious and deliberative cost-benefit maximization (or, on my account, preference formation and satisfaction) but the disposition to follow relevant social norms and expectations. People ‘let their actions be dictated by what we might call the cultural framing of the situation in which they find themselves’ (Pettit, 1995, p. 319). This ‘more or less automatic, cultural pilot’ (Pettit, 1995, p. 319) causes action in ways that sideline whatever deliberate considerations and judgments people might have. These have what Pettit calls a ‘virtual presence’. Cudd (2014, p. 36) analyses how ‘tacit commitments regulate behavior without being made conscious and explicit’. This fits my description of duties and commitments, which typically motivate people’s actions directly and can even have them choose against whatever comes out of their more conscious and deliberative preference formation processes (which imply comparing, evaluating and judging).

Not only Pettit’s ‘automatic, cultural pilot’ but also Searle’s and Sen’s committed people are essentially having their actions guided by norms. These are guidelines that tell us what (others empirically and normatively expect us) to do, depending on what (we empirically and normatively expect) others to do. Different circumstances highlight different norms and thus prescribe different kinds of behavior. When contextual cues trigger a norm (Binmore & Shaked, 2010, p. 98, Eckel & Gintis, 2010, p. 111), Hausman would claim that this makes people’s preferences – understood as total comparisons – conditional upon how people perceive a specific situation (see also: Bicchieri, 2006). In my view, and Pettit’s model captures this idea well, norms typically have a stronger and more direct impact on people’s behavior. They generally motivate behavior without involving reflection, deliberation and consideration (Cudd, 2014, pp. 50–53). Unless the costs of following the context-dependent norms become exceedingly high, people will not start comparing or considering whether or not to follow the rules of conduct at hand. In short, norms often bypass preferences instead of forming them.⑩
Sen (1985, p. 219; 2002, p. 217) himself already stressed this non-instrumental – and in my view potentially counterpreferential – character of rule-following. His notion of commitment allows him to understand people’s actions ‘as motivated by a rule that displaces the agent’s own goal.’ (Cudd, 2014, p. 48). As with duties – quite often properly understood as internalized moral norms – the idea is not that the resulting behavior is completely unreflective or heteronomous. Norm-guided actions are based on the reasons (we can see) we have to recognize and respect the people around us. In this respect, Davis (2007, p. 326) talks of ‘a rational recognition of rules associated with social membership that can only be achieved by (...) self-scrutinizing individuals’. This capacity to form commitments and bind oneself by the obligations and requirements coming from our social environment is not so much antithetical to personal agency as it is essential to it (Cudd, 2014, pp. 49–50; Davis, 2007, p. 331).

**Answering Hausman’s concerns**

Now, what is Hausman’s main concern with my definition of preferences as overall and thus partial comparative evaluations? The answer lies in the assumption of choice determination that he believes is crucially needed to link (rational) preferences to (consistent) choices.

Because preferences determine choices, they must take into account everything relevant to choice. If, as in everyday usage, Jill’s preferences leave out factors that influence her choices, such as moral commitment, then her preferences will not determine her choices. Choice determination implies that preference rankings are not partial rankings. (Hausman, 2012, p. 17)

This, however, puts the cart before the horse. Why can one not drop the assumption of choice determination altogether? If there are factors in preference-based models that determine choices besides preferences, namely beliefs and constraints, why not allow for more? My approach simply implies that BPC-models are extended to a BPDC-models (including also duties) or a BPNC-models (including also norms).

Another, related argument is that preferences can function well in game theoretical models and expected utility theory, only if they are all-things-considered rankings (Hausman, 2005, pp. 45–46). Lehtinen (2013, p. 209) summarizes Hausman’s point.

If one does not include all the motivating factors into the payoffs, the game is incompletely specified (...). In order to be able to concentrate on analysing the games, game theorists have good reasons to be as silent as possible on the content of the preferences.

While there is nothing wrong with this strategy, it clearly restricts the scope and explanatory force of such game theoretical models. While they can function as tools to understand the structure of different kinds of social interactions (which can take different forms when players are assumed to have different total comparative evaluations), they no longer help us understand why individuals come to act as they do.

Hausman admits this much. If the utilities present all-things-considered rankings, as he thinks they should, then ‘game theory has nothing to say about how individuals construct their all-things-considered rankings of comprehensive outcomes’ (Hausman, 2005, p. 47). Sure, Hausman (2005, p. 49) says, there are ‘intellectual costs of supposing that all the factors influencing choice can be captured by a single notion of preference’. In my opinion, these costs – in explanatory terms – are simply too big. If, as Hausman (2013, p. 219) himself admits, ‘tying preference so closely to choice
evades rather than solves the hard problems in explaining, predicting and evaluating choices’, I favor a strategy that encourages economists to tackle these problems.

**Going beyond semantics: modeling economic agents**

Now, one can ask whether the whole issue is anything more than ‘just semantics’ (Hausman 2005, p. 44). Does it matter how we define preferences and conceptually relate them to people’s actual motivations? I believe it does, because these discussions entail different views of what goes on in people’s minds and what causes people’s behavior. While Hausman (2012, p. 43) rightly argues that ‘the fact that final preferences are uninteresting (...) is no argument that they do not exist’, I believe there are good reasons to assume that they often do not exist, let alone be a major cause of action, in the cases analyzed before.\(^1\)

In fact, here I side with Hausman’s earlier doubts about the explanatory force of models based on all-things-considered preferences: ‘To attribute to people a consistently articulated all-things-considered ranking of alternatives is a strong idealization, which in many contexts may be extreme and unreasonable’ (Hausman, 2005, p. 43). While not going into details, he goes on to stress the importance of understanding ‘when such an idealization is sensible and when it is not’.

Now, it is exactly my point that talk of preferences as all-things-considered rankings paints a picture of agents that – even when understood as an idealization, an isolation or an ‘approximation’ (Hausman, 2012, p. 107) – often fails to capture what is going on. In the examples discussed, people do not form ‘cognitively demanding’ (Hausman, 2013, p. 219) all-things-considered judgments at all, because they simply do not ‘consider’ or ‘take into account’ all factors. What motivates these choices are factors that cannot be adequately understood as ‘evaluations’, ‘rankings’ or ‘comparisons’. If, as Hausman (2013, p. 253) states, ‘causal explanation requires that the cited causes and mechanisms exist and result in the phenomena to be explained and that our descriptions of the causes and mechanisms are approximately correct’, I believe that economists working with Hausman’s notion of preferences do not causally explain most dutiful and committed behavior, because their picture of the underlying causes and mechanisms – as I have fleshed them out – is not approximately true.

It is interesting to note that, in earlier work, Hausman agrees with my claim that it might not be reasonable to idealize human psychology in the way all-things-considered preferences do. Again, his case in point is morality.

It is hard to believe that the choices of actors who are influenced by moral concerns can be rationalized by a complete and transitive preference function, nor is it obvious that it is reasonable to demand that they can be. But even if utility maximization and morally motivated conduct are formally consistent, one might still question how helpful utility theory is in analyzing moral conduct. (Hausman & McPherson, 1993, pp. 687–688)

In my view, it is thus quite often not useful if one wants to explain individual actions. On a side note, Guala (2012, p. 145) rightly stresses that the claim ‘“I have a preference for chocolate” does not necessarily imply a commitment to the existence of a set of entities called preferences, any more than ‘Paul gave Helen a kiss’ implies a commitment to the existence of kisses’. According to Guala, preferences are not entities but, as one can expect when talking about ‘comparisons’ and ‘rankings’, relations between objects of choice. Nevertheless, like in the case of Paul and Helen kissing, an activity is claimed to occur when people prefer things: namely ranking alternatives and
considering all kinds of aspects. Again, I want to claim (1) that this activity often does not occur and (2) that when it occurs, it can be overrun by more direct influences on people’s choices.

**Conclusion**

Taking preferences to be *overall* comparative evaluations adequately captures the mental processes behind people’s choices. It forces economists and other behavioral scientists to take on the complex task of causally explaining human action and specify under which circumstances which mental factors enter into comparisons or when they influence action in more direct ways. My normative claims about what definition and role of preferences should have, should be of relevance to at least those economists who aim to explain human action. At least some of the models and theories that economists (and other behavioral scientists) build should ultimately be aimed at gaining knowledge about – among other things – human action.

Opening up the black box, economists should pay more attention to how people perceive choice situations and which of the available alternatives they take to be legitimate options. One example of this can be found in the literature on ‘crowding out’, where the introduction of financial incentives has been shown to undermine the force of intrinsic (often moral) motivations (Frey & Jegen, 2001). Take the day-care center where introducing fines for latecomers led to *more* parents showing up late (Gneezy & Rustichini, 2000). To understand this – to conventional economists surprising – phenomenon, one must see that the parents initially perceived their relationship toward the teacher as norm-guided, feeling they ‘should not take advantage of her patience’ (Gneezy & Rustichini, 2000, p. 14). Only after the fines were introduced, did arriving late become a legitimate and guilt-free option. Understanding preferences as overall comparative evaluations, I argue, allows for a more adequate causal explanation of all this: some parents always preferred arriving late but they perceived this as a viable alternative only after their sense of duty was crowded out. Their moral concern shifted from a ‘normative restraint’, causing them to act against their preference (before fines), to a consideration that they could weigh off against more pragmatic concerns (after fines).

Now, where does this approach depart from Sen’s? Sen’s primary strategy is to distinguish between the multiple conceptions of preferences in the literature without specifying which of these economists should employ: ‘I have no strong views on the correct use of the word “preference”’ (Sen, 1977, p. 329). Hausman (2005, pp. 34–35; 2012, pp. 20–22, pp. 63–65) rightly argues that Sen reveals the ambiguous usage of the term by economists and rational choice theorists but does not help to clear up the conceptual confusion this has generated. While I follow Hausman’s plea for a single and uniform definition of the term, I favor one that is narrower than his all-encompassing one. On my account, preferences encompass whatever goes into comparative judgments and evaluations but leave out those motivational factors that influence choice more directly and not via some comparative evaluation.

What would Sen himself think of my disagreement with Hausman? While he still does not want to endorse one single definition of preferences, he seems to have (almost) no problem with Hausman’s proposed definition of preferences as total comparative evaluations (Sen, 2007, pp. 353–354). Yet, he disagrees with Hausman on the possibility of an agent putting internal restraints on his own goal-seeking. Since Sen insists that this breaks the link between self-goal and choice, I believe he is defending the possibility of motivation-based counterpreferential choice. 12 Translating Sen’s talk
of ‘goals’ into the notion of ‘preferences’, this at least implies that preferences are but one factor motivating people’s actions. I thus follow Anderson (2001, p. 23) in interpreting Sen as proposing an ‘alternative model of behavior, commitment, that was not framed in terms of preference satisfaction at all’.

Even though Hausman and I are both critical of Sen’s approach, we agree with his fundamental claim that economics is in need of a more nuanced picture of real-life people’s motivations. One-dimensional *homo economicus* models that do not specify how the stipulated preferences are formed, where they come from and how they change, reinforce the tendency of economists to disregard human motivation and stress that actual people do not necessarily feel and think like the players in their models: ‘Neoclassical economists (…) follow Paul Samuelson in thinking it a virtue not to be committed to any particular psychological view of how human minds work.’ (Binmore & Shaked, 2010, p. 88). In my view, economists *should* be committed to particular psychological views of how human minds work. Without more empirically informed and conceptually sound substantive assumptions about people’s motivational and cognitive characteristics, economic models lack predictive and explanatory force.

If economists and other social and behavioral scientists want to explain not only the unintended consequences of people’s (inter)actions but also why people act as they do, they need to assess in what ways different factors influence action in different contexts. Only if they can construct a theory specifying which factors play which causal role in which circumstances, can positive economic models claim to causally explain people’s actions. In my view, understanding preferences as overall instead of total comparative evaluations is a crucial step in this direction.

**Acknowledgement**

Views expressed in this article have been presented on numerous occasions and I thank participants at these events for their useful comments. More specifically I want to thank the members of TINT (University of Helsinki, Finland), CEE (KU Leuven, Belgium), EIPE (Erasmus University Rotterdam, The Netherlands) and TiLPS (Tilburg University, The Netherlands). At the 3rd International Conference Economic Philosophy in Aix-en-Provence (June 2016), I benefited from insightful comments by, amongst others, Constanze Binder, John Davis, Wade Hands, Daniel Hausman and Aki Lehtinen. I want to thank them and two anonymous reviewers of this journal.

**Disclosure statement**

No potential conflict of interest was reported by the author.

**Notes**

1. This criticism on the usage of ‘preferences’ may be said to hold a fortiori for utility, ‘which is frequently taken to be an index of preference’ but which ‘has an even wider range of meanings’ (Hausman, 2005, p. 43). As William Stanley Jevons (1871, p. 31) already knew, a utility-based theory of action becomes tautological when the notions of ‘pleasure’ and ‘pain’ take on a broad meaning: ‘Call any motive which attracts to a certain action pleasure, and that which deters pain, and it becomes impossible to deny that all actions are prompted by pleasure or by pain’.

2. A related strategy has been to give up on causal explanation of micro-level explananda (individual action) and focus on macro-level explananda (e.g. the unintended consequences of human interaction).

3. I thank an anonymous reviewer of this journal for urging me to stress this distinction between the ‘clarification exercise’ and the ‘normative exercise’.
4. This definition is taken from the online Merriam-Webster Dictionary (http://www.merriam-webster.com/dictionary/motivation, accessed August 30, 2016).

5. As with the strategy of working with a broad conception of preferences, this claim that preferences are generated by duties and other motivations is not new. Fisher (1892/2007, p. 11) already argued that desires – or preferences – can be formed by different psychological factors: ‘No one ever denied that economic acts have the invariable antecedent, desire. Whether the necessary antecedent of desire is “pleasure” or whether independently of pleasure it may sometimes be “duty” or “fear” concerns a phenomenon in the second remove from the economic act of choice and is completely within the realm of psychology.’ While economists follow the ‘simple psycho-economic postulate: Each individual acts as he desires’, it is up to psychologists and others to study how desires come about.

6. It should be no surprise that Hausman (2005, pp. 47–48) makes room for counterpreferential choice in this sense of the word.

7. I am not claiming that duties always trump other considerations or are never weighed off against other motivations. In situations where people make conscious and deliberate trade-offs between duties (or commitments for that matter) and other considerations, one can adequately model them along the lines of Hausman. However, I believe that in most situations, they influence choices more directly. As will become clear later on (in the part on Pettit’s model of the virtual homo economicus), I believe that conscious evaluations, comparisons and trade-offs typically influence one’s choices in a virtual way rather than in an actual way.

8. Pauer-Studer (2007, p. 94) understands commitment as ‘an external (desire-independent) reason’. Like a moral principle, it can give people reason to do something that they do not desire. According to Cudd (2014), such an approach implies that something alien and externally to one’s motivations is causing one’s behavior, which means there is no genuine agency. Sen’s argument that commitment violates self-goal choice has met with the same criticism: no agency without self-goal choice (Schmid, 2005, p. 56). In my view, agency is more closely related to reasons than to desires. When people have good reasons to be committed, their agency is not undermined. In contrast with the standard notion of ‘autonomy-agency’, which requires self-goal choice, Cudd (2014, p. 50) proposes ‘a broader theory of human agency, in which (…) to be an agent is to act responsively in a normative framework’. Acting on duties, commitments, and (as we will see later on) internalized norms is one way of doing exactly that. One might wonder how such desire-independent reasons can ever motivate an agent. According to Searle (2001, p. 168), there are ‘secondary desires’ such as the desire to be a trustworthy person. While I take duties and commitment to be part of the agent’s motivations, I agree that higher-order desires play a role. Most people care about being a person that upholds basic norms of politeness. Again, this desire is not a factor in some all-things-considered comparison. When paying for a beer, one simply lives up to what one believes is important in life, not weighing this off against other aspects of the choice at hand.

9. Think of the (pre)commitment technique to remove a temptation. If you want to quit smoking but you fear being seduced by the craving for nicotine, you can avoid situations where you can buy and smoke cigarettes. Unlike Ulysses, who restricts himself physically by having him tied to the mast, you remain physically able to buy and consume cigarettes but your (pre)commitment to stop motivates you to not take the option that you know you will prefer.

10. While not only norms but also customs, habits, rules of thumb and psychological heuristics influence choices in this direct way, I have not focused here on such less intentional behavior. Of course, the idea dates back to Simon (1978, p. 362): ‘It is not that people do not go through the calculations (…) – neoclassical thought has never claimed that they did. What has been shown is that they do not even behave as if they had carried out those calculations.’ I do not claim that it is impossible for economists to model motivational factors such as commitments and duties as determinants of preferences and then stipulate that actual people act as if they were satisfying these preferences (in fact, Hausman is right that most economists have been doing this all along). I do claim that such as if models disregard the underlying human psychology and are thus missing a crucial element in adequate causal explanations of most committed and dutiful behavior.
11. I am not taking a stance in the debate about whether unobservables are real or not (Guala, 2012). Surely, while preferences are unobservable (Hausman, 1998, p. 196), there is no principled reason why they cannot be real and actually cause people’s actions. I disagree with Hausman not because I think preferences never exist (I believe they often do) but because I think they play a different role in the causal story behind people’s actions. In short, my quarrel does not concern the legitimacy of preference-based explanations but their scope.

12. In fact, Hausman (personal communication) agrees that there remains substantive disagreement between himself and Sen here (see also: Lehtinen, 2013, p. 209).

References

Anderson, E. (2001). Unstrapping the Straitjacket of ‘Preference’: A comment on Amartya Sen’s contributions to economics and philosophy. *Economics & Philosophy, 17*, 21–38.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge, MA: Cambridge University Press.

Binmore, K., & Shaked, A. (2010). Experimental economics: Where next? *Journal of Economic Behavior & Organization, 73*, 87–100.

Blaug, M. (1992). *The methodology of economics: How economists explain* (2nd ed.). Cambridge: Cambridge University Press.

Brennan, G. (2007). The grammar of rationality. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment* (pp. 105–123). Oxford: Oxford University Press.

Broome, J. (2001). Are intentions reasons? And how should we cope with incommensurable values? In C. Morris & A. Ripstein (Eds.), *Practical rationality and preference: Essays for David Gauthier* (pp. 98–120). Cambridge: Cambridge University Press.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117*, 817–869.

Croson, R., & Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization, 73*, 122–131.

Cudd, A. (2014). Commitment as motivation: Sen’s theory of agency and the explanation of behavior. *Economics & Philosophy, 30*, 35–56.

Davis, J. B. (2007). Identity and commitment: Sen’s fourth aspect of the self. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment* (pp. 313–335). Oxford: Oxford University Press.

Eckel, C., & Gintis, H. (2010). Blaming the messenger: Notes on the current state of experimental economics. *Journal of Economic Behavior & Organization, 73*, 109–119.

Fehr, E., & Gächter, S. (2002, January 10). Altruistic punishment in humans. *Nature, 415*, 137–140.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*, 817–868.

Fisher, I. (1892/2007). *Mathematical investigations in the theory of value and prices*. New York, NY: Cosimo.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys, 15*, 589–611.

Friedman, M. (1953). The methodology of positive economics. In M. Friedman (Ed.), *Essays in positive economics* (pp. 3–43). Chicago, IL: The University of Chicago Press.

Gintis, H. (2006). Behavioral ethics meets natural justice. *Politics, Philosophy & Economics, 5*, 5–32.

Gintis, H. (2009). *The bounds of reason: Game theory and the unification of the behavioral sciences*. Princeton: Princeton University Press.

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies, 29*, 1–17.

Guala, F. (2008). Experimental economics, history of. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave dictionary of economics online* (2nd ed.). New York, NY: Palgrave Macmillan. Retrieved September 1, 2016, from http://www.dictionaryofeconomics.com/article?id=pde2008_E000289

Guala, F. (2012). Are preferences for real? Choice theory, folk psychology, and the hard case for commonsensical realism. In A. Lehtinen & P. Ylikoski (Eds.), *Economics for real: Uskali Maki and the place of truth in economics* (pp. 137–155). London: Routledge.
Güth, W., & Kliemt, H. (2007). The rationality of rational fools: The role of commitments, persons and agents in rational choice modelling. In F. Peter & H. B. Schmid (Eds.), Rationality and commitment (pp. 124–149). Oxford: Oxford University Press.

Hausman, D. M. (1998). Problems with realism in economics. Economics & Philosophy, 14, 185–213.

Hausman, D. M. (2005). Sympathy, commitment, and preference. Economics & Philosophy, 21, 33–50.

Hausman, D. M. (2012). Preferences, value, choice, and welfare. Cambridge: Cambridge University Press.

Hausman, D. M. (2013). A reply to Lehtinen, Teschl and Pattanaik. Journal of Economic Methodology, 20, 219–223.

Hausman, D. M., & McPherson, M. S. (1993). Taking ethics seriously: Economics and contemporary moral philosophy. Journal of Economic Literature, 31, 671–731.

Hodgson, G. (2013). From pleasure machines to moral communities: An evolutionary economics without homo economicus. Chicago, IL: University of Chicago Press.

Jevons, W. S. (1871). The theory of political economy. London: Macmillan and Co.

Lehtinen, A. (2013). Preferences as total subjective comparative evaluations. Journal of Economic Methodology, 20, 206–210.

Lipsey, R. G. (2008). Positive economics. In S. N. Durlauf & L. E. Blume (Eds.), The New Palgrave dictionary of economics (2nd ed.). New York, NY: Palgrave Macmillan. Retrieved September 1, 2016, from http://www.dictionaryofeconomics.com/article?id=pde2008_P000130

Mäki, U. (1992). On the method of isolation in economics. In C. Dilworth (Ed.), Idealization IV: Intelligibility in science. special issue of poznan studies in the philosophy of the sciences and the humanities (Vol. 26, pp. 317–351). Amsterdam: Rodopi.

Mäki, U. (2009). Missing the world: Models as isolations and credible surrogate systems. Erkenntnis, 70, 29–43.

Mansbridge, J. (1998). Starting with nothing: on the impossibility of grounding norms solely in self-interest. In A. Ben-Ner & L. Putterman (Eds.), Economics, values, and organization (pp. 151–168). Cambridge: Cambridge University Press.

Mill, J. S. (1836/1874). On the definition of political economy; and on the method of investigating proper to it. In Essays on some unsettled questions of political economy. London: Longmans, Green, Reader, and Dyer. Retrieved August 31, 2016, from http://www.econlib.org/library/Mill/mllUQP5.html

Pattanaik, P. K. (2013). The concepts of choice and preference in economics. Journal of Economic Methodology, 20, 215–218.

Pauer-Studer, H. (2007). Instrumental rationality versus practical reason: Desires, ends, and commitment’. In F. Peter & H. B. Schmid (Eds.), Rationality and commitment (pp. 49–69). Oxford: Oxford University Press.

Pettit, P. (1995). The virtual reality of ‘Homo Economicus’. Monist, 78, 308–329.

Pettit, P. (2005). Construing Sen on commitment. Economics & Philosophy, 21, 15–32.

Rabin, M. (1993). Incorporating fairness into game theory and economics. The American Economic Review, 83, 1281–1302.

Reiss, J. (2012). The explanation paradox. Journal of Economic Methodology, 19, 43–62.

Schelling, T. C. (1969). Models of segregation. American Economic Review, 59, 488–493.

Schmid, H. B. (2005). Beyond self-goal choice: Amartya Sen’s analysis of the structure of commitment and the role of shared desires. Economics & Philosophy, 21, 51–63.

Schmidt, D. (1993). Reasons for altruism. Social Philosophy and Policy, 10, 52–68.

Searle, J. R. (2001). Rationality in action. Cambridge, MA: MIT Press.

Sen, A. (1977). Rational fools: A critique of the behavioral foundations of economic theory. Philosophy & Public Affairs, 6, 317–344.

Sen, A. (1985). Goals, commitment and identity. Journal of Law, Economics and Organization, 1, 341–355.

Sen, A. (2002). Rationality and freedom. Cambridge, MA: Harvard University Press.

Sen, A. (2005). Why exactly is commitment important for rationality? Economics & Philosophy, 21, 5–14.

Sen, A. (2007). Rational choice: Discipline, brand name, and substance. In F. Peter & H. B. Schmid (Eds.), Rationality and commitment (pp. 339–361). Oxford: Oxford University Press.
Simon, H. (1978). Rational decision-making in business organization. In A. Lindbeck (Ed.), *Nobel lectures: Economics 1969–1980* (pp. 343–371). Singapore: World Scientific Publishing Co.

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, 7, 1–31.

Walras, L. (1874/2003). *Elements of pure economics*. (W. Jaffé, Trans.). Routledge Library Editions: Economics. London: Routledge.