



Under Risk, Over Time, Regarding Other People: Language and Rationality Within Three Dimensions

Dorian Jullien

► To cite this version:

Dorian Jullien. Under Risk, Over Time, Regarding Other People: Language and Rationality Within Three Dimensions. Research in the History of Economic Thought and Methodology, 2018, 36 (C). halshs-01651042

HAL Id: halshs-01651042

<https://shs.hal.science/halshs-01651042>

Submitted on 28 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Under Risk, Over Time, Regarding Other People: Language and Rationality Within Three Dimensions

Dorian Jullien*

September 1, 2017

Abstract

This paper conducts a systematic comparison of behavioral economics's challenges to the standard accounts of economic behaviors within three dimensions: under risk, over time and regarding other people. A new perspective on two underlying methodological issues, i.e., interdisciplinarity and the positive/normative distinction, is proposed by following the entanglement thesis of Hilary Putnam, Vivian Walsh and Amartya Sen. This thesis holds that facts, values and conventions have interdependent meanings in science which can be understood by scrutinizing formal and ordinary language uses. The goal is to provide a broad and self-contained picture of how behavioral economics is changing the mainstream of economics.

Keywords: behavioral economics, economic rationality, expected utility, prospect theory, exponential discounting, hyperbolic discounting, self-interest, other-regarding behaviors, economic methodology, history of economics, philosophy of economics, economics and language

JEL: A12, B21, B41, D01, D03, D81, D90, D64

Introduction

The intended contribution of this paper is to provide a broad picture of how behavioral economics is changing the mainstream of our discipline. Behavioral economics stems from the

*Université Côte d'Azur, CNRS, GREDEG, France. dorian.jullien@gredeg.cnrs.fr. I thank Samuel Ferey, Peter Wakker, Nicolas Vallois, Wade Hands, Kevin Hoover and two anonymous referees for their comments on earlier drafts of this paper, along with the participants to the 2015 Thurstone Workshop in Paris. This paper is a revision of a part of my Ph.D. dissertation (chap.2, sect. 1) and a few developments on risk preferences in section 3 are also used in a previously published paper (in *Economia*, 6-2, 2016, 265-291).

influence, in the late 1970s, of psychologists Daniel Kahneman and Amos Tversky on economist Richard Thaler. Since then, behavioral economists borrow psychologists’ experimental methods and theoretical concepts to study various empirical departures of individual behaviors from the predictions of standard models. Most contributions in behavioral economics focus on three dimensions of economic rationality, which are often expressed differently depending on one’s audience. Addressing economists, Stefano DellaVigna uses the language of preferences: “[t]he first class of deviations from the standard model [...] is nonstandard preferences [...] on three dimensions: time preferences, risk preferences, and social preferences” (2009, p.316). Addressing scholars from other disciplines, Thaler and his co-authors stress “Three Bounds of Human Nature” (Mullainathan and Thaler 2001, p.1095), namely that “people exhibit bounded rationality, bounded self-interest, and bounded willpower” (Jolls, Sunstein and Thaler 1998, p.1471). In both cases, these authors are saying that behavioral economics mainly challenges standard models of individual behaviors under risk, over time and regarding other people.¹

Alternative models from behavioral economics can be seen as modifications – conceptually inspired from psychology – to standard models. The most popular ones are reprinted in *Advances in Behavioral Economics* (Camerer, Loewenstein and Rabin 2004). There again, a large part of the “basic topic” is organized in terms of “preferences over risky and uncertain outcomes”, “intertemporal choice” and “fairness and social preferences” (ibid, table of content). This threefold distinction is not new. It organizes numerous classics of standard economics, albeit in a less salient fashion (see e.g., Deaton and Muellbauer 1980; Mas-Colell et al. 1995). It is, however, increasingly used in a systematic way by prominent standard economists writing on behavioral economics (e.g., Fudenberg 2006; Pesendorfer 2006; Levine 2012). Despite the large number of contributions from historians, philosophers and methodologists on issues related to behavioral economics, there is no reflexive perspective on the centrality of this threefold distinction. This paper conducts a historically-motivated and methodologically-oriented comparison of the three ways behavioral economics shaped itself *within* each dimension.

The historical motivation is to give an account of the empirical and theoretical state of affairs

¹In terms of Sent’s (2004) distinction, this paper is only about the “new” behavioral economics of Kahneman, Tversky, Thaler and others but not the “old” one initiated by Simon. Connections and contrasts between these two traditions are discussed elsewhere (see, e.g., Sent 2004; Klaes and Sent 2005, esp. p.47; Egidi 2012; Kao and Velupillai 2015; Geiger 2016). On the systematic deviations that are not about the three dimensions but contradict directly the bedrock of consumer choice theory under certainty (in the sense of Mas Collé et al. 1995, chaps. 1-3), see Thaler (1980; 2015, parts. I-II).

around behavioral economics before the second half of the 2000s. Before this period, behavioral and standard economists focused on either one of the three dimensions taken separately from the other two. After this period, the focus is *also* on what happens when at least two dimensions are interacting. In other words, there is a shift of focus from contributions *within* the three dimensions to contributions *across* the three dimensions. This shift is arguably of historical significance because it produces opportunities for a reconciliation between standard and behavioral economics. Contributions across the three dimensions are however too numerous and disorganized to be discussed here. The historical scope of this paper is conceptual, not chronological, i.e., it also uses post mid-2000s contributions *within* the three dimensions to better understand the conditions of possibility of the contemporary shift – not the shift itself.²

The methodological orientation is to provide a new perspective on two interrelated issues underlying the rise of behavioral economics. On the one hand, there is the positive/normative issue, which is about the meaning of economic rationality in the interplay between (positive) observations of individual behavior and (normative) judgments of these observations as rational or not. On the other hand, there is the issue of interdisciplinarity, which is about the way behavioral economists borrow from psychologists to modify standard economics's models. The new perspective on these two issues consists in scrutinizing the role of language uses by both economists and economic agents, especially about the three dimensions of risk, time and other people and their associated psychological terms. The first section of the paper presents this methodological perspective in some details, which is then used throughout the paper. The second one compares the main empirical challenges that have motivated behavioral economists' work on the three dimensions. The third one discusses the common denominator to these challenges, namely that they all violate instances of the consequentialism inherent to standard models. The last one compares behavioral economics's main theoretical alternatives in the three dimensions. Finally, a conclusion briefly summarizes the arguments and suggests further researches of historical, methodological and philosophical interests.

There are three limits to the broadness of the picture presented here due to space constraints. Firstly, it focuses much more on the *qualitative* challenges from behavioral economics than on

²Representatives of the shift to contributions across dimensions include Abdellaoui et al. (2013), Andersen et al. (2008), Andreoni and Sprenger (2012), Baucells and Heukamp (2012), Bolton and Ockenfels (2010), Epper and Fehr-Duda (2015), Halevy (2008), Kovarik (2009), Saito (2013), Trautmann and Vieider (2012).

the *quantitative* ones. Qualitative challenges are inconsistent behaviors that violate some axioms underlying the standard models, making it impossible to represent those behaviors by *one* utility function. Quantitative challenges are behaviors that can be captured by one such function but at the price of implying absurd numerical values and degrees of curvature (interpreted as *measures* of risk, time *and/or* social preferences in different situations, see Broome 1991). Secondly, it focuses on decision theoretic issues at the expense of game theoretic ones. Hence the dimension of other people is investigated in its simplest forms. Thirdly, it focuses on probabilistic risk at the expense of non-probabilistic uncertainty.³

1 Methodological Perspective

Both behavioral economists and commentators in history, methodology and philosophy of economics remark that the rise of behavioral economics comes with a use of the positive/normative distinction that is different from the traditional one (see, e.g., Thaler 1980; Heukelom 2011; 2012; 2014; Hands 2015). The traditional use of the distinction is to separate a positive economics describing, explaining or predicting economic states of affairs from a normative economics evaluating, recommending or prescribing economic states of affairs (Mongin 2006a). Here ‘normative’ takes an ethical meaning: economists make or study ethical value judgments, i.e., about what is good, fair, just and so on. Furthermore, the main objects of discussion are collective forms of behavior or rationality and the evaluation of market efficiency. The different use of the distinction associated with the rise of behavioral economics is to separate a positive from a normative interpretation of models of individual behavior. On a positive interpretation, such models describe, explain or predict the behavior of an individual. On a normative interpretation, they evaluate, recommend or prescribe *rational* behavior. Hence, ‘normative’ here does not (at least directly) take an ethical meaning: economists make or study value judgments of rationality, i.e., what one ought to do to be rational, what counts as rational and so on. By and large, behavioral economists disagree with a positive interpretation of standard models of individual behavior,

³Game-theoretic issues involve some inherent connections with the other two dimensions (e.g., the role of ‘strategic uncertainty’ in ‘repeated games’, see Camerer 2003; though see Aumann and Dreze 2009 on the connection between game and decision theory). On decision making regarding what POSSIBLY happens, though not probabilistically so, see the literature on ambiguity (e.g., Wakker 2010). On what may be thought of as UNREAL, see the literature on ignorance and unawareness (e.g., Zeckhauser 2014 provides some references and claims that this is where the future of the economics of uncertainty lies to understand the real world).

i.e., they do not describe, explain and predict individual behavior well enough. Therefore, behavioral economists propose alternatives or modifications to standard models allowing for a positive interpretation. However, behavioral economists, by and large, agree with a normative interpretation of standard models, i.e., they allow to evaluate, recommend or prescribe rational behaviors. In other words, standard models justify value judgments of rationality about behaviors that conform to these models – and accordingly may justify value judgments of irrationality about behaviors that do not conform to these models⁴.

One methodological presupposition underlying both uses of the positive/normative distinction is the fact/value dichotomy, i.e., that the (positive) observation of facts is in principle independent of (normative) values (Mongin 2006b). This presupposition has been expressed and contested in many ways throughout the last century in philosophy (see, e.g., Marchetti and Marchetti 2017) as well as in economics (see, e.g., Blaug 1998; Mongin 2006b; Castro Caldas and Neves 2012; Reiss 2017). The methodological orientation taken in this paper follows one existing criticism of the fact/value dichotomy: the entanglement thesis of Hilary Putnam, Vivian Walsh and Amartya Sen (see Putnam and Walsh 2011). One specificity of the entanglement thesis consists in bringing together arguments against the fact/value dichotomy with arguments against another methodological presupposition: that the (positive) observation of facts is in principle independent of social or theoretical conventions, i.e., the fact/convention dichotomy. The two sets of arguments that Putnam brings together in a mutually-reinforcing way can be summarized as follows (for more details, see Putnam 2002; 2004; 2013; 2015). The arguments against the fact/convention dichotomy come from two directions. From the philosophy of science, they are the arguments about the predictive and explanatory role of unobservable theoretical entities in the natural sciences. From the philosophy of logic and of mathematics, they are the arguments about the indispensability of these disciplines in the natural sciences and about their widely

⁴Expressions such as ‘value judgments of rationality or irrationality’ are shortcuts for the following observation which is common to most methodological and philosophical discussions of rationality in decision theory: judgments of violations of decision-theoretic axioms as ‘rational’ or ‘irrational’ do not come from the structure of the theory but from the values of the scientist(s) (e.g., decision theorists, experimental economists using decision theory). It can be argued that conflicting values regarding what counts as ‘rational’ behavior sparked much of the philosophical and methodological debates over the meaning of rationality in decision theory (which might not have occurred if scholars only spoke in terms of ‘violations of decision-theoretic axioms’ and never added the words ‘rational’ or ‘irrational’). This observation is in line with Mongin’s remark that “[i]n the hands of many economists, individual decision theory obeys a [...] pattern. [...] [T]hey would *define* an individual agent to be someone who maximizes an objective under an availability constraint, and then normally *add* the value judgment that optimization is an essential part of rationality” (2006b p.260).

recognized objectivity without making reference to observable objects. The arguments against the fact/value dichotomy also come from two directions. From pragmatist philosophers, they are the arguments about the necessity of *epistemic* value judgments, e.g., ‘simplicity’, ‘relevance’, and so on, for the construction and selection of scientific theories. From meta-ethics, they are the arguments about the existence of predicates that are inseparably positive and normative, e.g., ‘cruel’, ‘rude’, by contrast to purely normative (in the ethical sense) predicates, e.g., ‘good’, ‘right’.

Putnam applies his work to economics mostly through Walsh’s historical and methodological work (see Walsh 1996; Putnam 2002). Sen has been taken by both Putnam and Walsh as an illustration of how economic theory could be practiced self-consciously without the fact/value and fact/convention dichotomies (see esp. Sen 1987; 2002; for an explicit agreement, see Sen 2005; 2009, esp. the fn p.357). In these contributions, another specificity of the entanglement thesis is a careful attention paid to the use of language in economics. More precisely, Putnam, Sen and Walsh pay a specific attention to the interplay between the formal language of mathematics, logic and probability theory and the ordinary language shared with economic agents about economic life. They scrutinize this interplay to account for the triple entanglement of facts, values and conventions in economics. Indeed, their criticisms of the two dichotomies do not imply that they are not useful distinctions to understand scientific practice. It only implies that facts, values and conventions do not acquire their meaning independently of each others. In other words, the entanglement thesis pays specific attention to language uses in order to understand how facts, values and conventions acquire meaning interdependently. The goal of this paper is to scrutinize this triple entanglement in behavioral economics – a domain not studied by Putnam, Sen and Walsh – to give a new perspective on the positive/normative issue it raises for models of individual behavior. It should be noted that ‘conventions’ is used in a broad way by Putnam Sen and Walsh, encompassing the use of mathematics, logic and probability theory in scientific theories, along with methodological conventions and social conventions (usually to show that these entities are entangled with facts and values hence not purely conventional). In this paper I will mainly use ‘theoretical conventions’ to speak of mathematical properties of economic models.

Another central methodological issue in behavioral economics is the issue of interdisciplinary,

which is about the way behavioral economists borrow from psychologists to modify standard economics's models (see, e.g., Sent 2004; Heukelom 2014; Hausman 2011, part III). This issue can be tackled through the entanglement thesis, and hence connected to the positive/normative issue, by investigating *the interactions* between the uses of two different types of psychological terms. There are, on the one hand, the folk or ordinary psychological terms – e.g., ‘preferences’ or ‘impatience’ – that economists use in their models as much as economic agents in their daily lives. On the other hand, there are the technical or scientific psychological terms used by behavioral economists as an interdisciplinary borrowing from psychology – e.g., ‘loss aversion’ or ‘impulsivity’. One task of the comparative exercise of this paper is to scrutinize the interactions between these two different uses of psychological terms *about the three dimensions*. To do so in a systematic and meaningful way, I will pay specific attention to the way these terms articulate the distinctions made in Table 1, which comes from the work of linguist-semiotician François Rastier (1996, fig. 14.2; see also 2008, p.213; 2012, fn41 p.23). Rastier remarks that, in all the different human languages studied by linguists, there is a common set of distinctions marked by different terms that we use to talk about our environment. Rastier argues that marking and articulating these distinctions is an anthropologically fundamental characteristic – because of the similarities in both what are the relevantly distinct dimensions of the environment and what are the relevantly distinct zones within each dimension. In a way this observation is corroborated by the centrality of the distinction between the three dimensions (of risk, time and people) in economics. Furthermore, we shall see that most of the challenges posed by behavioral economics for a given dimension involves marking a distinction between at least two zones of that dimension. Rastier capitalizes the words in the table because they represent specific types of information that can be instantiated in communication through a variety of other words, e.g., ‘SURE’, ‘LIKELY’, ‘IMMEDIATELY’, ‘LATER’, ‘ME’, ‘US’, ‘DEAR READER’, etc. Throughout this paper, information about the three dimensions are also capitalize to make them explicit when they are implicit.⁵

Many commentators have already pointed out that the relation between economics and psy-

⁵Two precisions are worth making. Firstly, Table 1 modifies Rastier's original tables by calling ‘(Un)certainity’ what he calls “*Mode*”, translating “*Personnes*” in French as ‘People’ in English and removing his fourth dimension, ‘Space’ (with HERE, THERE and OVER THERE/ELSEWHERE as its three zones), because it is not a relevant dimension of individual behavior in behavioral economics. Secondly, the perspective on interdisciplinarity taken here is consistent with, and partly inspired by, historian of psychology Danziger (1997) and, to a lesser extent, Heukelom's (2009) use of Galison (1999).

	Identity Zone	Proximal Zone	Distal Zone
(Un)certainty	CERTAIN	PROBABLE	POSSIBLE, UNREAL
Time	NOW	ONCE, SOON	PAST, FUTURE
People	I, WE	YOU	HE, IT

Table 1: Similarities of language uses across the three dimensions

chology at play in the development of behavioral economics shows some general inconsistencies with the fact/value and fact/convention dichotomies – without of course using this terminology. For instance, Angner and Lowenstein (2012) emphasize the role played by unobservable theoretical entities about the psychology of individuals in the observation of behaviors that are judged irrational from standard models. Heukelom (2014) shows how behavioral economics stems from psychologists using standard economic models first as set of theoretical conventions for empirical measurements and then also as a set of normative values to judge behaviors as irrational. Angner (2015) interprets the role of standard models in behavioral economics as ideal types playing both analytic and normative functions. Grüne-Yanoff (2015) focuses on the dimension of time to show that after a short period of interdisciplinary exchanges between economics and psychology, behavioral economics became interested in empirical phenomena different from the ones of interests to psychologists, mostly because of different theoretical and normative concerns. Heidl (2016) characterizes behavioral economics as a de-idealization of standard economic models through empirical observations and theoretical considerations inspired from psychology which have sometimes – though rarely – suggested norms of rationality different than standard models (see p.82). In a sense, all these accounts presuppose the interdependence of facts, values and conventions at play in the entanglement thesis. Therefore, the methodological perspective of this paper is twofold. On the one hand, it consists in making explicit the entanglement of facts, values and conventions in the classical contributions from behavioral economics within the three dimensions. On the other hand, it consists in highlighting structural similarities – and sometimes implicit interactions – across the three dimensions that are at play in these classical contributions. Finally, the account of behavioral economics presented below tries to strike a better balance between (broadness by taking the three dimensions altogether) and precision (by entering as much as possible into empirical and theoretical details) than one can find in other existing accounts.

2 The main empirical regularities, altogether and naked

Drazen Prelec and George Loewenstein (1991) propose a systematic comparison of the dimensions of risk and time in behavioral economics. The exposition presented here follows theirs to a great extent. The main differences are that (1) the three dimensions are taken *altogether*, i.e., adding the dimension of other people and (2) empirical patterns of preferences are presented *naked*, i.e., with as less theoretical, methodological and normative considerations as possible.⁶

One of the most robust behavioral phenomena in the dimension of risk is what Kahneman and Tversky (1979; problems 3 and 4) coined as the *certainty effect* – a specific instance of the so-called *common ratio effect*. The decision maker faces two choices. The first one is between a CERTAIN consequence (here \$3000) and PROBABLE consequences (here \$4000 and \$0), and the observed pattern is usually:⁷

$$\text{The CERTAINTY of winning \$3000} \succ 80\% \text{ CHANCE of winning \$4000}$$

The certainty effect is the conjunction of this pattern with the one obtained when all the above consequences remain constant but their probabilities are reduced by a common factor (here by four), i.e., the choice between two sets of PROBABLE consequences:

$$25\% \text{ CHANCE of winning \$3000} \prec 20\% \text{ CHANCE of winning \$4000}$$

There is a strong similarity between the certainty effect in the dimension of risk and the *immediacy effect* in the dimension of time – a specific instance of the so-called *common difference effect*. A first choice is between consequences occurring NOW and consequences occurring LATER (or IN THE FUTURE) and a second one is between the same consequences all delayed LATER or IN THE FUTURE by the same amount of time, i.e., none occur NOW anymore. Here are two illustrations of this pattern. The first one is a non-monetary intuitive example that has somewhat the status of an institution in behavioral economics (from Thaler 1981, opening example):

$$\begin{aligned} 1 \text{ apple TODAY} &\succ 2 \text{ apples TOMORROW} \\ 1 \text{ apple IN A YEAR} &\prec 2 \text{ apples IN A YEAR AND A DAY} \end{aligned}$$

⁶Also, for ease of exposition, all currencies have been converted into dollars, efforts have been made to put the reader into an experimental subjects' shoes and extreme instances of the regularities have been chosen to make their 'logic' more salient – abstracting from issues of replications and effect size (e.g., whether 51% or 100% of the subjects of an experiment display a given behavioral phenomenon).

⁷As usual, ' \succ ' means 'is (strictly) preferred to'.

The second is a monetary one useful for the present comparative exercise (from Faralla et al. 2012, stimuli 107 and 109):

$$\begin{array}{lll} \text{Winning \$5 TODAY} & \succ & \text{Winning \$7.50 in A MONTH} \\ \text{Winning \$5 in TWO WEEKS} & \prec & \text{Winning \$7.50 in A MONTH AND TWO WEEKS} \end{array}$$

To include the dimension of other people with as less strategic interaction as possible in this comparison, consider the original *dictator game* from Kahneman, Knetsch and Thaler (1986, experiment 2, part 1). The decision maker faces only one choice, i.e., dividing a sum of money (here \$20) between HIMSELF and ANOTHER anonymous person, and the observed pattern is usually:⁸

$$\text{\$18 to SELF and \$2 to OTHER} \prec \text{\$10 to SELF and \$10 to OTHER}$$

The phenomena illustrated so far are the basic ones in behavioral economics. Conjunctions of their underlying patterns of preferences with further patterns of preferences constitute further empirical challenges known as *sign effects* and *magnitude effects* (Prelec and Loewenstein 1991, pp. 774-5) – that are empirically more complex and theoretically more problematic than the basic ones. A well-known sign effect under risk is the conjunction of the certainty effect with the reversed pattern of preferences occurring when the signs of monetary consequences are simply reversed from gains to losses (from Kahneman and Tversky 1979, problems 3' and 4'):

$$\begin{array}{lll} \text{The CERTAINTY of losing \$3000} & \prec & \text{80\% CHANCE of losing \$4000} \\ \text{25\% CHANCE of losing \$3000} & \succ & \text{20\% CHANCE of losing \$4000} \end{array}$$

To illustrate a sign effect over time, take the conjunction of the immediacy effect involving monetary gains with the reversed pattern of preferences occurring when gains become losses (from Faralla et al. 2012, stimuli 107 and 109):

$$\begin{array}{lll} \text{Losing \$5 TODAY} & \prec & \text{Losing \$7.50 in A MONTH} \\ \text{Losing \$5 in TWO WEEKS} & \succ & \text{Losing \$7.50 in A MONTH AND TWO WEEKS} \end{array}$$

Magnitude effects over time consist in keeping delays constant and increasing the amounts of money (i.e., the *magnitude* of the consequence) by a common factor (here by six), which triggers no immediacy effects within the gain domain and within the loss domain but an overall sign effect (from Faralla et al. 2012, stimuli 117 and 119):

⁸Ultimatum games, trust games and so on involve too much strategic interaction for the scope of this paper.

Winning \$30 TODAY	\prec	Winning \$45 in A MONTH
Winning \$30 in TWO WEEKS	\prec	Winning \$45 in A MONTH AND TWO WEEKS
Losing \$30 TODAY	\succ	Losing \$45 in A MONTH
Losing \$30 in TWO WEEKS	\succ	Losing \$45 in A MONTH AND TWO WEEKS

Magnitude effects under risk consist in keeping (un)certainty constant and increasing the amounts of money by a common factor (here by ten), which triggers no certainty effects for small gains, certainty effects for small losses and another type of sign effect than the one seen previously (from Markowitz 1952, pp.153-2):

The CERTAINTY of winning 10 cents	\prec	10% CHANCE of winning \$1
The CERTAINTY of winning \$1,000,000	\succ	10% CHANCE of winning \$10,000,000
The CERTAINTY of losing 10 cents	\succ	10% CHANCE of losing \$1
The CERTAINTY of losing \$1,000,000	\prec	10% CHANCE of losing \$10,000,000

By contrast with the dimensions of risk and time, in the dimension of other people as studied in dictator games, it is *the absence* of magnitude effects that is taken as an empirical challenge (see, e.g., Levitt and List 2007, p.164; and the references in Novakova and Flegr 2013). In one of the rare study displaying a magnitude effect, Julie Novakova and Jaroslav Flegr (2013, Table 1) strongly increase the amount of (hypothetical) money to be allocated: four times by a factor of 10 – e.g., \$20, \$200, \$2000, \$20000, \$200000. The proportion of the allocation given to THE OTHER reduces constantly from 28.3% to 23.3%. Even in that case, the allocations to THE OTHER are well within the 20-30% range of aggregate results on standard dictator game, with less than 40% of decision maker giving nothing (see Camerer 2003, chap.2; Engel 2011).⁹

Though the parallel with the decisions' consequences is easy because it is identical in experiments on the three dimensions, i.e., money, the parallel between the decisions' dimensions is less straightforward, i.e., between risk or time, on the one hand, and other people, on the other hand. The notion of *social distance* provides one plausible parallel. A traditional dictator game always involves the three zones in the dimension of people: a decision maker (I, THE ALLOCATOR) communicating to the experimenter (YOU, THE ECONOMIST) how much to give

⁹By contrast with the original dictator game from Kahneman et al. (1986), most dictator games do not present a decision problem between two allocations but can allocate a given amount of money between HIMSELF and THE OTHER as he wishes, i.e., he/she can choose within a quasi-continuous distribution of monetary consequences.

to someone else (HE, THE RECIPIENT) who is usually anonymous. In standard designs, there is close to zero communication between the allocator and the recipient; though both communicate with the experimenter. Roughly, social distance in a dictator game characterizes the communication among these three protagonists, whether and how they talk to each other and what they know about each others' identity and actions. The challenging empirical regularities are that variations of social distance can push allocators *both ways*: either to give more than 20-30%, or, at the opposite, to keep more than 80-70% (see Engel 2011, sect. 4.5). Two classical papers illustrate this point. Notably using double blind experimental design where both the experimenter and the recipient do not know who gives to whom, Elizabeth Hoffman, Kevin McCabe and Vernon Smith (1996) show how the less communication between the allocator and the experimenter, the more the former is likely to keep 100% of the money for himself. Keeping communication between the allocator and the experimenter very low, Iris Bohnet and Bruno Frey (1999) show how the more communication between the allocator and the recipient (from the former just seeing the latter, to both seeing each others, through the former seeing and hearing personal information about the latter), the more the former is likely to propose above 20-30% of the allocation to the latter. *The identity of the recipient* is indeed one of the rare variables that triggers (mean) allocation where the allocator keeps less than 50% for himself, as when the recipient is from a poor community in a third world country (Brañas-Garza 2006).¹⁰

Sign effects are not thoroughly studied in dictator games. The few contributions on the topic start by endowing both the allocator and the recipient with money (List 2007; Bardsley 2008). The allocator is allowed to share the money with, as well as to take money from, the recipient. That tends to push the allocator to give 0%, but not to take from the recipient. Such experiments do not (nor were they intended to) parallel the studies of losses under risk or over time where decision problems involve no positive monetary consequences at all. Therefore potential sign effects in dictator game are yet to be investigated (though see Antinyan 2014 in economics; and Leliveld et al. 2009, exp.3, in social psychology).

Here are three conclusions one can draw from comparing behavioral economics's empirical challenges in the three dimensions altogether. Firstly, these behavioral phenomena involve, in

¹⁰'Social distance' was also used long ago by Edgeworth to discuss what is today called social preferences (see Fontaine 2000, p.413), and is currently used by behavioral psychologists making the case that it is the interpersonal analogue of probability and time in decision making (see e.g., Jones and Rachlin 2006; and Ida and Ogawa 2013 for a discussion in economics).

terms of Rastier’s table, interactions of different zones *within* a given dimension. Secondly, each dimension involve implicit assumptions: that the SURE or PROBABILISTIC consequences under (un)certainty are IMMEDIATE and for ONESELF (i.e., not DELAYED and/or for SOMEONE ELSE), that the IMMEDIATE or DELAYED consequences over time are SURE and for ONESELF, and that the consequences of an allocation between ONESELF and ANOTHER are SURE and IMMEDIATE. Thirdly, dictator games point a more general implicit assumption: that (at least ‘lab’) experiments always involve more than the identity zone in the dimension of people (I, WE) as the experimenter addresses the decision maker and/or *vice-versa* by using words belonging to the proximal zone in the dimension of people (YOU).

3 Violations of consequentialism

This section compares the theoretical implications of the above empirical challenges for standard models of risk, time and social preferences. A notational and terminological note is helpful at this point. An object of choice X can be represented as a set of consequences $\{x_1, x_2, \dots, x_n\}$ distributed either probabilistically (p_1, p_2, \dots, p_n summing to 1), or over time (t_1, t_2, \dots), or across people (i, j, k, \dots). Objects of choice are called *lotteries* when x_1 has p_1 chance to occur (and so on) so that $X = (x_1, p_1; \dots; x_n, p_n)$; or *plans* when x_1 happens at t_1 (and so on) so that $X = (x_1, t_1; x_2, t_2; \dots)$; or *allocations* when x_1 is for individual i (and so on) so that $X = (x_1, i; x_2, j; \dots)$. Theoretical conventions of standard models ensure the equivalence of choice, preference and utility. A decision maker chooses an object of choice if and only if this object is preferred to the others in the choice set, and if and only if more utility is derived from it than from the others in the choice set. In formal language, with $C(\cdot)$ denoting a choice function, \succsim a preference relation and $U(\cdot)$ a utility function, theoretical conventions ensure the conditions under which $C(\{X, Y\}) = X \Leftrightarrow X \succsim Y \Leftrightarrow U(X) \geq U(Y)$.

Standard models of risk preferences are built within the theoretical framework of *expected utility*. The expected utility ($U(\cdot)$) of a lottery is a real number given by the sum of the evaluations of its consequences by a function ($u(\cdot)$), linearly weighted by their respective PROBABILITIES. A formal illustration of the expected utility of ‘80% CHANCE of winning \$4000’ in the certainty effect is $U(X) = u(x_1)p_1 + u(x_2)p_2 = .8u(\$4000) + .2u(\$0)$. The absence of a function

that could change the numerical values of the probabilities represents the linear weighting of expected utility. The meaning of $u(\cdot)$ and its relation with $U(\cdot)$ under certainty are matters of controversy not discussed here. To avoid making controversial statements, $u(\cdot)$ is called a “*subutility function*” (following Gorman 1968, p.368) that ‘subevaluates’ the consequences or give real numbers representing the ‘subutilities’ of the consequences, while $U(\cdot)$ evaluates lotteries.¹¹

The crucial theoretical convention ensuring the equivalence of choice, preference and expected utility is the so-called *independence axiom*. It states that a preference for a lottery over another should not change if a third lottery is mixed with both in the same way. For instance, $pX + (1 - p)Z$ is the mix of lottery X with lottery Z , and this mix can be weighted by any probability p . The independence axiom requires that a preference between two lotteries $X \succsim Y$ does not change when both are mixed with Z with the same weight: $pX + (1 - p)Z \succsim pY + (1 - p)Z$ (i.e., the mix with X is still preferred to the mix with Y). Hence the independence axiom is often expressed as $X \succsim Y \Leftrightarrow pX + (1 - p)Z \succsim pY + (1 - p)Z$. In the certainty effect, $X = (\$3000, 1)$, $Y = (\$4000, .8; \$0, .2)$, with pattern of preference $X \succ Y$, and Z is the certainty of getting nothing, i.e., (with abuse of ordinary language) the lottery $(z_1, p_1) = (0, 1)$. The objects of choice involved in the second pattern of preference, i.e., ‘25% chance of winning \$3000 \prec 20% chance of winning \$4000’ are mixes with equal weights $p = .25$, $1 - p = .75$ of X and Y with Z , i.e., $.25X + .75Z = (\$3000, .25; \$0, .75)$ and $.25Y + .75Z = (\$4000, .2; \$0, .8)$. Hence the two patterns of preferences violate the independence axiom; magnitude and sign effects under risk also violate the independence axiom.¹²

Standard models of time preferences are built within the theoretical framework of *exponentially discounted utility*. The discounted utility ($U(\cdot)$) of a plan is a real number given by a weighted sum of the subutilities of the consequences. Unlike the linear weighting by probabilities in expected utility, the weighting by time here is exponential through a discount function ($\delta(\cdot)$). From the period $t - 1$, the subutilities of the consequences that happen at t

¹¹Roughly, the issue is that $u(\cdot)$ has cardinal properties for comparison among consequences that $U(\cdot)$ does not need for comparison among lotteries (or among temporal plans or social allocations for the two other dimensions, see below). If both are strongly related, then one can infer cardinal utility in decision under certainty, *contra* the ordinalist revolution (see esp. Ellsberg 1954; Fishburn 1989; Wakker 1994; Ellingsen 1994; Guala 2000; Abdellaoui et al. 2007; Mongin 2009; Moscati 2013a; b; 2016a; b). This gets even more problematic when one gives a psychological (*contra* purely technical) meaning to the cardinality of $u(\cdot)$, as is the case in behavioral economics (see Rabin 2000, fn3 p.1282; Wakker 2010).

¹²There are other ways to state the independence axiom formally and informally (see MacCrimmon and Larson 1979; Fishburn and Wakker 1995; and Mongin 2009); it can also be decomposed in sub-parts (see Machina 1989; and Burghart et al. 2015).

are discounted by $\delta^{t-1} = (\frac{1}{1+\rho})^t$, where ρ is the *constant* rate of discounting that the decision maker uses indifferently in all periods. To illustrate, in Thaler’s apple example with $t_1 = \text{today} = 1$, $t_2 = \text{tomorrow} = 2$, $W = (1 \text{ apple}, t_1; 0 \text{ apple}, t_2)$ and $X = (0 \text{ apple}, t_1; 2 \text{ apples}, t_2)$: $U(W) = \delta^{1-1}u(1 \text{ apple}) + \delta^{2-1}u(0 \text{ apple}) = u(1 \text{ apple}) + \frac{u(0 \text{ apple})}{1+\rho}$ and $U(X) = \delta^{1-1}u(0 \text{ apple}) + \delta^{2-1}u(2 \text{ apples}) = u(0 \text{ apple}) + \frac{u(2 \text{ apples})}{1+\rho}$. Hence δ^0 is always equal to 1, so that the subutility of the first consequence is never discounted, that of the second one is discounted by $1 + \rho$, that of a third one would have been discounted by $(1 + \rho)^2$ (though the discounting between the second and third periods is still $1 + \rho$), and so on exponentially. The parameter ρ is often called *pure time preference* and represents the decision maker’s *patience* (when it tends to 0 or becomes negative) or *impatience* (when it grows). Hence standard models of time preferences introduce folk-psychological terms related to the dimension of time (i.e., patience and impatience), while, in a sense, linear weighting by probability in standard models of risk preferences prevent such introduction.

The crucial theoretical convention in the framework of exponential discounting is called *dynamic consistency*, which is an implication of axioms – not an axiom *per se* as is the independence axiom in expected utility. Dynamic consistency ensures the consistent evaluation of the same plan through time, in the sense that its discounted value does not depend on the period from which it is evaluated. In the apple example, the first pattern of preference ‘1 apple today \succ 2 apples tomorrow’ should hold if and only if $U(W) > U(X)$. Denoting the time of evaluation by a superscript (U^t), dynamic consistency further ensures that $U^1(W) = U^{365}(W)$ and $U^1(X) = U^{365}(X)$ so that $U^1(W) > U^1(X) \Leftrightarrow U^{365}(W) > U^{365}(X)$. The same reasoning should hold for the second pattern of preference ‘1 apple in a year \prec 2 apples in a year and a day’: with $Y = (1 \text{ apple}, t_{365}; 0 \text{ apple}, t_{366})$ and $Z = (0 \text{ apple}, t_{365}; 2 \text{ apples}, t_{366})$, $U^1(Y) < U^1(Z) \Leftrightarrow U^{365}(Y) < U^{365}(Z)$. However, it is easy to see that the evaluation of Y and Z in a year is identical to the evaluation of W and X today: $U^{365}(Y) = U^1(W)$ and $U^{365}(Z) = U^1(X)$. Hence $U^1(W) > U^1(X)$ and $U^{365}(Y) < U^{365}(Z)$ violate dynamic consistency; magnitude and sign effects over time also violate dynamic consistency.¹³

¹³More precisely, the two main axioms implying dynamic consistency are ‘separability’, which has the same mathematical structure as the independence axiom and gives EDU its separably additive form, and especially ‘stationarity’, which requires the subutility of a given consequence to be the same for any period; the latter is obviously central in yielding dynamic consistency (see Fishburn and Rubinstein 1982; Bleichrodt et al. 2008; Lapiard and Renault 2012).

By contrast with the linear weighting by probability and the exponential weighting by time, other people should not, in principle, be weighted in a specific way within standard models of social preferences. However, in practice, decision makers are often assumed to be driven by their self-interest. What this means in terms of weighting other people is less clear than the weighting in the other two dimensions. This is so because standard models of social preferences do not rest on a formal axiomatic framework in the same sense in which standard models of risk and time preferences do. In other words, folk-psychological terms about other people are not imposed by standard models of social preferences, they stem from their interpretation in ordinary language. As Walsh puts it, “the assumption of self-interest is not presented as a formal axiom – it does not come in, as it were, by the front door” (Walsh 1996, p.113). Nevertheless, systematic discussions of self-interest are possible through Sen’s (2002, chap.1, sects 8-10) characterization of various ways by which self-interest enters standard models in economic theory, as it were, by the back door. Sen argues that “three requirements” are usually “imposed in the traditional models” (ibid, p.34):

“*Self-centered welfare*: A person’s welfare depends only on her own consumption and other features of the richness of her life (without any sympathy or antipathy towards others [my emphasis], and without any *procedural concern* [my emphasis]).
Self-welfare goal: A person’s only goal is to maximize her own welfare.
Self-goal choice: A person’s choices must be based entirely on the pursuit of her own goals” (ibid, pp.33-4)

In dictator games, only self-centered welfare is systematically and straightforwardly violated. These violations can be interpreted in a notation related to the ones used so far, with the standard though controversial identification of ‘welfare’ and ‘utility’ – thus the requirement can just be called *self-centeredness*. The utility of an allocation evaluated by ONESELF (i) is a real number ($U^i(\cdot)$) given by the sum of the subutilities of its consequences for ONESELF ($u_i(\cdot)$) and OTHER PEOPLE ($u_{-i}(\cdot)$, with $-i = j, k \dots \neq i$). Self-centeredness requires that for any consequence x , $u_{-i}(x) = 0$ (see Sobel 2005, sect.3.1). In Kahneman et al.’s dictator game, with $i = \text{self}$, $j = \text{other}$, $X = (\$18, i; \$2, j)$ and $Y = (\$10, i; \$10, j)$, self-centeredness implies $U^i(X) = u_i(\$18) + u_j(\$2) = u_i(\$18) > U^i(Y) = u_i(\$10) + u_j(\$10) = u_i(\$10)$, which is violated by the preference pattern $Y \succ X$. Indeed, any dictator game where the decision maker allocates something to the other – even the tiniest amount of money – violates self-centeredness. However,

that violates only the “consumption” and “others” parts of the requirement of self-centeredness (italicized in the above quote). Pretty much all other variations in experimental design that trigger variations in allocations violate the “procedural concern” part of the requirement. It is not difficult to see that behaviors in dictator games do not necessarily violate self-welfare goal and self-goal choice.¹⁴

The independence axiom, dynamic consistency and self-centeredness are all justified by value judgments derived from consequentialism, i.e., the principle that the consequences of a choice are the only sources of reasons from which that choice can be justified. Consequentialism is explicit only in self-centeredness through its requirement that “procedural concern” should not influence an decision makers’ choices. It is made formally explicit in Peter Hammond’s work on standard models in the three dimensions (Hammond 1976; 1977; 1983; 1987; 1988a; b; 1989; 1998; Hammond and Zank 2014). Roughly, Hammond formally defines consequentialism as an axiomatic requirement on choice functions to show how it *implies* dynamic consistency, the independence axiom and self-centeredness. In his contributions, dynamic consistency becomes the normative benchmark to evaluate risk and social preferences, even in atemporal settings (i.e., unreduced decision trees for decision under risk and *ex ante* and *ex post* comparison of social allocations). In a sense, this conclusion is not surprising. Of the three dimensions, time is the one where standard models impose the stronger necessary folk-psychological constraints, i.e., a constant degree of patience (or impatience). By contrast, standard models of risk preferences do not allow decision makers to be especially optimistic or pessimistic about the probability that a given consequence occurs; and standard models of social preferences do not necessarily require that decision makers are narrowly self-centered.¹⁵

Consequentialism involves further folk-psychological terms about the three dimensions when the quantitative issues underlying the sub-evaluations of consequences by subutility functions are considered. Depending on research goals or subfields, an interpretation of the curvature of $u(\cdot)$ as the marginal utility of consequences can allow the inference of different unobservable behavioral information. The three prominent are, informally stated, measures of decision makers’ attitudes

¹⁴More precisely, self-centeredness does not imply the violation illustrated here by itself, the other standard assumption of monotonicity (usually made in all standard models) is also necessary here. It requires that utility increases with the quantity of consequences.

¹⁵Peter Wakker pointed to me that, historically, the first contribution to anticipate and understand Hammond’s conclusion is from philosopher Burks (1977, chap.5) (personal communication 20/04/2016).

towards more or less equal distributions of consequences over (1) different probabilistic states of the world, (2) different points in time and (3) different people (see Broome 1991). In other words, there are different interpretations of the same mathematical construction. When $u(\cdot)$ is concave, (1) and (3) are interpreted through the technical though folk-psychologically-inspired terms of *risk aversion* and *inequality aversion*. On the other hand, whatever the curvature, (2) remains interpreted through the purely technical terms of *elasticity of intertemporal substitution*.¹⁶

Both Sen and Hammond emphasize that consequentialism is one of the deepest source of value judgments in economics (with Hammond defending consequentialism as a sound and potentially universal norm of rationality and Sen highlighting its limits). The origin of this is, according to them, the philosophical influence of utilitarianism in economics (see also Walsh 1996). Another influence can be noted. As logicians Dov Gabbay and John Woods put it: “[h]istorically, [...] logic is an examination of consequentialist reasoning whose success or failure is definable for or representable in semi-interpreted formal languages” (2005, p.21). Hence, the widespread acceptance of consequentialist reasoning in standard economics can also be attributed to the epistemic value judgments favoring the uses of formal languages in economic theory, especially through the axiomatic method. In the words of the entanglement thesis, the *factual* patterns of choice that violate the theoretical *conventions* of the independence axiom, dynamic consistency and self-centeredness are subject to *value* judgments from consequentialism. Put simply, criticisms of these patterns of choice as irrational are often made explicitly or implicitly from consequentialism; and defenses of these patterns of choice as rational often involve criticisms of (at least some features of) consequentialism. Be they critical or defensive, these discussions involve references to both Rastier’s distinctions within, and folk psychological terms about, a given dimension. The next section scrutinizes how these methodological issues

¹⁶Three strategies from economists to deal with this can be identified. Firstly, most economists just focus on one of the three dimensions while explicitly abstracting from one of the two others (see esp. Strotz 1956, p.166, focusing on time abstracting from risk or Pratt 1964, p.123, focusing on risk with no specific role for time). Secondly, some economists think that there are good reasons to separate the dimensions and have accordingly specified more general functional forms to give different measures to two of the three different attitudes (see esp. Kreps and Porteus 1978 on risk and time preferences). Thirdly, other economists search normative justifications for why the three types of attitudes should be identical, and in empirical cases where they are not, which one should play the normative benchmark defining how one of the other two should change (see esp. Harsanyi 1955; 1988 on risk preferences as a benchmark for social preferences). These three strategies usually focus on two dimensions and are totally silent on a third one (for more representative examples, see also Gollier 2001; Gajdos and Weymark 2012; as far as I know, the only contributions about the three dimensions altogether are Güth et al. 2008; Atkinson et al. 2009; and to a lesser extent Broome 1991). Because of their bearing on quantitative challenges and contributions across the three dimensions, these issues are outside the scope of this paper.

arise in the main theoretical alternatives from behavioral economics to deal with the empirical challenges.¹⁷

4 Theoretical alternatives from behavioral economics

This section scrutinizes the entanglement of facts, values and theoretical conventions in alternatives from behavioral economics, first under risk, then over time and finally regarding other people.

Under risk

Prospect theory is behavioral economics' main theoretical alternative to expected utility theory. The difference between its two main versions – Kahneman and Tversky's (1979) original prospect theory and Tversky and Kahneman's (1992) cumulative prospect theory – underlies non-trivial aspects of the issue of interdisciplinarity and the positive/normative issue in behavioral economics. In the original prospect theory, the utility of a lottery $X = (x_1, p_1; x_2, p_2)$ is a real number given by $U(X) = u(x_1)\pi(p_1) + u(x_2)\pi(p_2)$. By contrast with expected utility, probabilities can weight consequences nonlinearly through a function $\pi(\cdot)$, the *weighting function*. The issue of interdisciplinarity can be seen, following Peter Wakker (2010, p.162), as Kahneman and Tversky (1979) bringing psychologists' probabilistic sensitivity (i.e., $\pi(\cdot)$ without $u(\cdot)$) together with economists' consequentialist sensitivity (i.e., $u(\cdot)$ without $\pi(\cdot)$). Put simply, the weighting function transforms a given probability p in the *decision weights* $\pi(p)$ that the decision maker puts on the utility of consequences. Exactly what shape the weighting function takes is an empirical and quantitative issue, where decision weights do not necessarily sum up to 1.¹⁸

Figure 1 modifies Kahneman and Tversky's (1979) original graphical representation of the weighting function to illustrate the transformation of probability in decision weight for the

¹⁷For a concise statement of the central issues underlying 'logical consequences', see Cozic's (2009) presentation of a classical text from Tarski (2009 [1936]) and the text itself. Also, economics arguably sneaked in the rationality debates in psychology and cognitive sciences (also related to Kahneman and Tversky's work) through the door of consequentialism (see the economic examples and reasoning used by Baron 1994 in a debate within psychology).

¹⁸That decision weights do not necessarily sum up to one makes them formally different from additive (i.e., standard) subjective probabilities but similar to non-additive subjective probabilities. Furthermore, decision weights can be interpreted either as the irrational misperception of objective probabilities or as the rational over- or under-weighting of the importance of some consequences. Both additive and non-additive subjective probabilities may be interpreted as irrational misperception (e.g., due to lack of information) but only non-additive subjective probabilities can be interpreted as over- or under-weighting (though there are other possible interpretations).

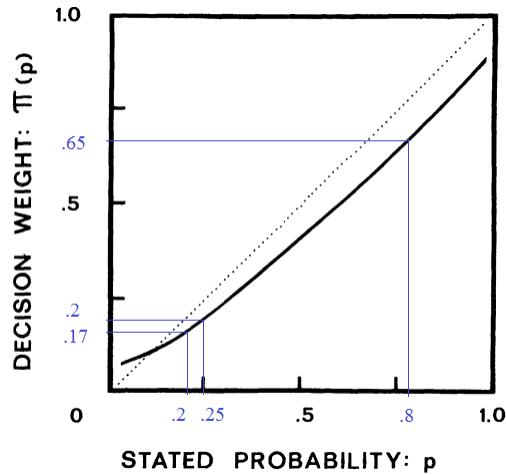


Figure 1: Kahneman and Tversky's 1979 Weighting Function (modified) and the Certainty Effect

certainty effect. In prospect theory, ‘the certainty of \$3000 \succ 80% chance of winning \$4000’ is equivalent to $u(\$3000) > .65u(\$4000)$ and ‘25% chance of winning \$3000 \prec 20% chance of winning \$4000’ to $.2u(\$3000) < .17u(\$4000)$. This is one instance of how prospect theory can formalize violations of the independence axiom. Note how the (1979) weighting function is normalized at the end point, i.e., $\pi(0) = 0$ and $\pi(1) = 1$, but not well-behaved near the end points 0 and 1. On Kahneman and Tversky's (1979, p.282) account, this represents two “categorical distinction[s]”: between impossibility and uncertainty near the end point 0, i.e., within Rastier's distal zone of (un)certainty, and between uncertainty and certainty and uncertainty near the end point 1, i.e., between Rastier's identity zone and the other two zones of (un)certainty.

The main problem with this weighting function is that it allows to formally represent preferences against first-order stochastic dominance. First-order stochastic dominance is a monotonicity condition under risk. Put informally, preference for first-order stochastic dominance ensures that lotteries with some chance to win some money are preferred to lotteries with less chance to win less money. Economists working in decision theory usually consider it as “a hallmark of rationality” (Wakker 2010, p.65), and its absence as “a fatal flaw sufficient to damn theories” (Bardsley et al. 2010, p.130). In other words, preference for first-order stochastic dominance is the most important theoretical convention in decision theory from which economists derive value

judgments of rationality and irrationality. This is why prospect theory has a serious normative drawback from an economist's perspective.¹⁹

Kahneman and Tversky (1979, p.274) try to bypass this problem by distinguishing two phases of decision making. The *evaluation phase* corresponds to the evaluation of lotteries through decision weights (i.e., with the weighting function). The *edition phase* corresponds to the process leading to the evaluation phase, represented as a set of psychological operations. One of these operations, the *detection of dominance*, "involves the scanning of offered prospects to detect dominated alternatives, which are rejected without further evaluation" (ibid, pp.274-5). This informal way of preventing a formal implication never convinced standard economists. What convinced at the same time standard economists, behavioral economists and even Tversky and Kahneman in their (1992) cumulative prospect theory is the so-called rank-dependent utility approach pioneered by economist John Quiggin (1982), for three main reasons. First, it formally modifies psychologists' decision weights to avoid violations of first-order stochastic dominance while still accounting for empirical challenges such as the certainty effect. Second, it shares expected utility's axioms with a weakened form of independence (see Quiggin 1982; Quiggin and Wakker 1994). Third, it allows folk-psychological explanations in terms of *optimism* and *pessimism*, which admit normative interpretations (see Diecidue and Wakker 2001).²⁰

The main distinction between standard models with rank-dependent utility and cumulative prospect theory is that Tversky and Kahneman (1992) use two rank-dependent weighting func-

¹⁹Systematic violations of first-order stochastic dominance comes from the non-additivity of decision weights, especially for small probabilities, e.g., $\pi(.01) + \pi(.06) > \pi(.07)$, so that '1% chance of winning \$10, 6% chance of winning \$10 and 97% chance of winning nothing' can be preferred to '7% chance of winning \$11 and 97 chance of winning nothing' (for more detailed explanations see Kahneman and Tversky themselves 1979, pp.283-4; see also Wakker 2010, sect. 5.3).

²⁰Quiggin's decision weights cannot be meaningfully illustrated with the certainty effect because its specificity shows up for lotteries that have at least three non-null consequences. It does not apply to consequences' probabilities, but to ranks, i.e., to cumulative probabilities. Hence consequences need to be ranked, i.e., the subscripts of the x s are meaningful, e.g., $x_1 \geq \dots \geq x_n$. Decision weights for a given consequence are derived from the entire probability distribution of consequences. More precisely, they are derived from the *ranks* of consequences x_i , which are the probability of receiving a given consequence *or* anything better, i.e., the *cumulative probability* $p_i + \dots + p_n$, not just p_i . Take the lottery (borrowed from Wakker 2010, pp.158-9) where you have $\frac{1}{6}$ chance of winning \$80, $\frac{1}{2}$ chance of winning \$30, $\frac{1}{3}$ chance of winning \$20, i.e., $X = (x_1, p_1; x_2, p_2; x_3, p_3) = (\$80, \frac{1}{6}; \$30, \frac{1}{2}; \$20, \frac{1}{3})$. There are four ranks: 0, $\frac{1}{6}$, $\frac{2}{3}$ and 1, corresponding respectively to the cumulative probabilities of more than \$80 (0), \$30 ($0 + \frac{1}{6}$), \$20 ($0 + \frac{1}{6} + \frac{1}{2}$) and \$0 ($0 + \frac{1}{6} + \frac{1}{2} + \frac{1}{3}$). Probabilities of individual consequences can be expressed in terms of *differences between ranks*: the probability of 'winning \$80' is $\frac{1}{6}$ which is equal to the rank of winning more than \$30 ($\frac{1}{6}$) minus the rank of winning more than \$80 (0); by the same reasoning, the probability of 'winning \$30' is $\frac{1}{2}$ (i.e., $\frac{2}{3} - \frac{1}{6}$) and the probability of 'winning \$20' is $\frac{1}{3}$ (i.e., $1 - \frac{2}{3}$). Finally, the decision weight w_i of a consequence x_i is the difference between ranks transformed by a weighting function $\pi(\cdot)$ such that $U(X) = u(x_1)w_1 + u(x_2)w_2 + u(x_3)w_3 = u(\$80) \times [\pi(\frac{1}{6}) - \pi(0)] + u(\$30) \times [\pi(\frac{2}{3}) - \pi(\frac{1}{6})] + u(\$20) \times [\pi(1) - \pi(\frac{2}{3})]$.

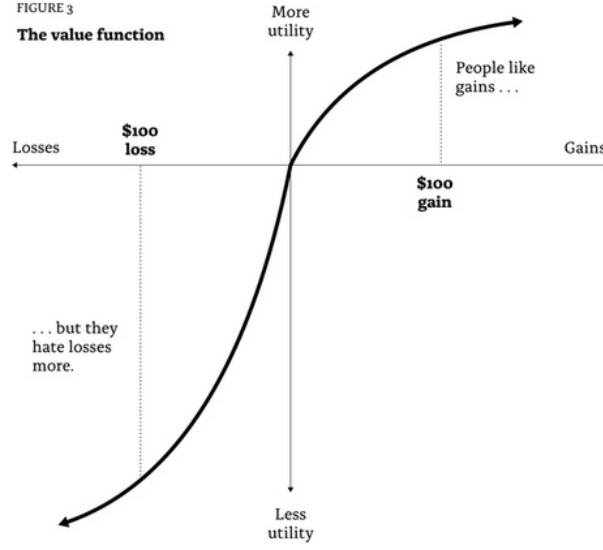


Figure 2: Prospect Theory's Value Function from Thaler (2015, fig.3)

tions (i.e., extending Quiggin's original approach) for gains and losses. Hence it avoids the normatively unpleasant violations of stochastic dominance while explaining more empirical phenomena than both standard rank-dependent models and original prospect theory (see Fennema and Wakker 1997; or Wakker 2010 appendix 9.8 for comparisons). The distinction between gains and losses is indeed central in behavioral economics and underlies its main conflicts with standard economics. If rank-dependent utility makes a step towards psychologists' probabilistic sensitivity, it does not adopt Kahneman and Tversky's specific perspective on consequences. Both the original (1979) and cumulative (1992) prospect theory make an explicit and central distinction between gains and losses by using the same *value function* $u(\cdot)$. It has a specific and empirically-motivated functional form that is pedagogically illustrated by Thaler (2015, fig.3); see Figure 2.

This function allows to represent *reference-dependent preferences* by including a *reference point* where the two axes cross. Its curvature is S-shaped around the reference point so that utility is concave for gains, convex for losses, and asymmetrically so to represent *loss aversion*: a difference in magnitude between utility from gains and disutility from losses (represented by the different lengths of the dotted lines). Interpreting the reference point as zero or as a given endowment allows prospect theory to account straightforwardly for various sign effects.

Reference-dependence allows to represent consequences taken as gains or losses with respect to a reference point: $u(x)$ or $u(-x)$. Standard economists' way of dealing with asymmetries between gains and losses (with either expected utility or rank-dependent utility models) through *asset integration* is less straightforward. It consists in representing consequences *on* the decision maker's final wealth through its initial wealth w , not consequences *per se*: $u(w+x)$ or $u(w-x)$. The most salient clash between these two approaches arises with respect to a quantitative challenge, the so-called Rabin paradox. Behavioral economists Matthew Rabin (2000) considers the following preference:

$$\text{nothing} \succ 50\% \text{ CHANCE of winning \$110 and } 50\% \text{ CHANCE of losing \$100}$$

Through some algebraic manipulations, he shows that absurd quantitative measures of risk aversion are implied by expected utility with asset integration, regardless of both initial wealth levels (and functional forms except for concavity, i.e., risk aversion). In particular, the previous preference for nothing implies the following one:

$$\text{nothing} \succ 50\% \text{ CHANCE of losing \$1000 and } 50\% \text{ CHANCE of winning any sum of money}$$

In short, plausible risk aversion over win/lose small stakes imply implausible risk aversion over large ones (e.g., replace 'any sum of money' by \$1.000.000). Rabin's solution through prospect theory's loss aversion raise at least three interrelated issues worth discussing here.²¹

First, as for probability weighting, loss aversion is not necessarily irrational if a decision maker consciously puts twice more weights on losses than on gains (Wakker 2010, p.239). This underlies the most frequent criticism of prospect theory: there is only sporadic considerations but no systematic theory of how the reference point changes, which has to be guessed in a pragmatic way by decision theorists (see, e.g., Wilkinson and Klaes 2012, p.190; Wakker, 2010, p.241-2 and sect.8.7; Stommel 2013). Thus, there is a lack of systematicity not only about the

²¹Rabin's results come from the whole of risk attitudes being derived from the marginal utility of consequences in expected utility, i.e., from economists' lack of probabilistic sensitivity. For more on the Rabin paradox, see Rabin and Thaler (2001), Barberis, Huang and Thaler (2006); Wakker (2010, pp.244-5) discusses most of standard economists' reactions to the paradox. There is a tradition in standard economics (esp. in experimental economics) to model *expected utility over income* rather than over final wealth. Many responses to the Rabin paradox use this strategy but without remarking that this is tantamount to reference-dependence (see Harrison, Lau and Rutström's 2007 appendix; and Wakker 2010, p.245). Because Rabin's solution is through loss aversion which entails reference-dependence, the only remaining disagreement between users of expected utility over income and prospect theory seems to be about the (ir)rationality of decision makers.

fixation of the reference point, but also, by implication, on the underlying (ir)rationality of the behavior under analysis.

Second, one way of resolving the asset integration *versus* asset isolation conflict is to drop neither and, instead, to make value judgments about the content of decision makers' preferences. This is Wakker's position (2010, chap.8), who argues that rationality requires risk neutrality (or very weak risk aversion) over small consequences. In other words, Wakker argues against the translation of consumer sovereignty in decision under risk. In other words, he argues that the content of preferences (for or against risk) is amenable to rational criticism.

Third, the real difference between standard analysis and prospect theoretic analysis is not the notion of reference-point *per se*. It is rather that a reference point is fixed in standard analysis while it can change in prospect theory. Indeed, expected utility's asset integration and prospect theory's reference-dependence are in agreement when prospect theory's reference point is a fixed level of initial wealth, so that there is "a unique relation between the [consequences] and the final wealth" (Wakker 2010, p.238). This unique relation is simply initial wealth plus consequences equals final wealth. It is when, for a given decision maker and a given set of choices, this unique relation is broken *during the analysis* by a change in reference point that both approaches disagree. In such cases, the traditional behavioral economics position of keeping the standard approach as the normative benchmark and using psychologists' theories as positive models surfaces clearly: "traditional EU [expected utility] is, in my opinion, the hallmark of rationality, any deviation from final wealth due to reference dependence is utterly irrational" (Wakker 2010, p.245). In short, "[e]xpected utility is the right way to make decisions" (Thaler 2015, p.30).²²

²²Among the attempts at endogenizing changes of reference point (e.g., Schmidt 2003; Schmidt, Starmer and Sugden 2008), Köszegi and Rabin (2006, 2007, 2009) can be mentioned as specifically relevant for the object of the paper because their account involves the three dimensions at once. Their determination of reference-point is (1) counterfactual (see e.g., 2006, fn3 p.1138), (2) temporal and (3) it uses tools from game theory (to model interpersonal relations between the selves of a decision maker). Furthermore, regarding the positive/normative issue, the determination is rational, i.e., they assume rational beliefs (i.e., rational expectation) for simplicity and often stress (esp. in the 2006 and 2007 conclusions) the inability of their models to account for some irrational behaviors demonstrated in behavioral economics (though see 2009, sect. III, where some of these irrational behaviors are seen as rational through the lenses of their model).

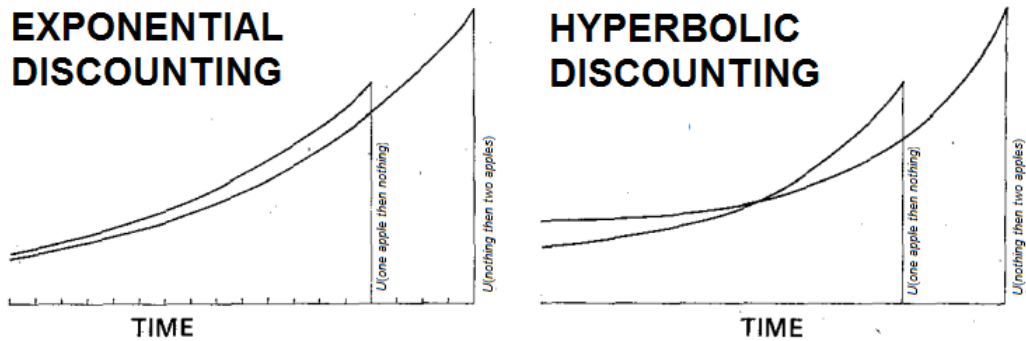


Figure 3: Exponential *versus* Hyperbolic Discounting and the Apple Example (modification to Ainslie 1975, fig.1)

Over time

Psychiatrist George Ainslie (1975; 1992) and his theory of hyperbolic discounting plays a role for time preferences in behavioral economics comparable to Kahneman and Tversky's prospect theory for risk preferences. Ainslie put together a large amount of empirical observations and theoretical considerations from various disciplines on human and animal behaviors over time. He makes the case for the importance of two behavioral phenomena: *impulsiveness*, responsible for dynamic inconsistencies and *impulse control*, i.e., various ways by which humans and animals try to avoid dynamic inconsistencies. Hence, the divide between standard and behavioral economics in the dimension of time is primarily marked by the distinction between impatience and impulsivity.

Figure 3 illustrates this with Thaler's (1981) apple example. Intertemporal preference reversals, i.e., dynamic inconsistencies, are impossible under exponential discounting because the rate of discounting, i.e., the degree of impatience, is constant. Hence, a choice for 'one apple early and then nothing' should imply that its utility is always superior to the utility of 'nothing and then two apples later'. This contradicts the outcome of Thaler's example. Under hyperbolic discounting, however, the utility of 'nothing and then two apples later' is superior to the utility of 'one apple early and then nothing' at early evaluation times, before smoothly reversing as the time of getting the early apple gets closer. The closer evaluation time from the occurrence of the

consequence, the greater the rate of discounting. Hence the rate of discounting is non-constant. Ainslie interprets this as impatience *and impulsiveness*.

The most famous introduction of Ainslie's work in economics was made by behavioral economist David Laibson (1994; 1997) in his models of *quasi-hyperbolic discounting*. His models bring together hyperbolic and exponential discounting, at least to some extents. The quasi-hyperbolic discount function in Laibson's models is a step function. It discounts the utility of a plan as its exponentially discounted utility with an additional discounting parameter β . This parameter is (i) necessarily equal to one at the time of evaluation and (ii) can then take any value between zero and one to multiply all exponentially discounted subutilities. Formally, quasi-hyperbolic discounting is represented by $U^t(X) = u(x_t) + \beta \sum_{i=1}^{\infty} \delta^i u(x_{t+i})$. It is easy to see that if $\beta = 1$, then quasi-hyperbolic discounting reduces to exponential discounting and implies dynamic consistency. It is also easy to see that if $\beta = 0$, then the decision maker does not care at all about LATER and THE FUTURE in the sense that all consequences after the date of evaluation contribute nothing to the utility of a plan under evaluation. In this case the utility of a plan is just the subutility of its INSTANTANEOUS consequence. In-between, i.e., for $0 < \beta < 1$, the closer β gets to 0, the greater the evaluation of a plan is biased by the PRESENT consequence (with respect to an evaluation under exponential discounting). Hence the two alternative labels for quasi-hyperbolic preferences in the literature, *present-biased preferences* and β - δ preferences. Like hyperbolic discounting, quasi-hyperbolic discounting can imply the preference reversals underlying dynamic inconsistency. Unlike hyperbolic discounting, the reversal is not smooth with quasi-hyperbolic discounting. Before the time of evaluation coincides with the occurrence of consequences, quasi-hyperbolic discounting is like exponential discounting but 'restrained', as it were, by $\beta < 1$. When the time of evaluation coincides with the occurrence of consequences, then $\beta = 1$, which suddenly makes their utility higher, especially compared to the utility of consequences occurring later.

Figure 4 provides a graphical representation of this explanation. If time is discrete, then kinks due to β suddenly becoming equal to 1 represent the beginning of the evaluation period

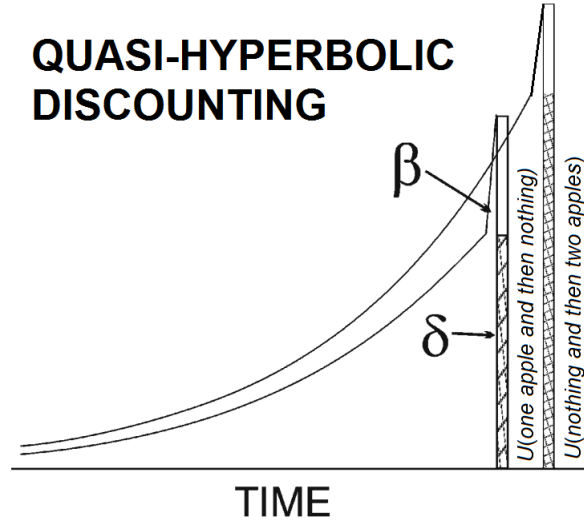


Figure 4: Quasi-hyperbolic Discounting and the Apple Example (modification to Ainslie’s 2012, fig.3)

when consequences (either one or two apples) occur.

The origin of the quasi-hyperbolic functional form is tightly related to the centrality of the three dimensions in economics. It comes from Edmund Phelps’s and Robert Pollak’s (1968) model intergenerational of altruism in macroeconomics. In their model, the PRESENT generation is denoted by t , and is evaluating a plan where the consequences are its OWN consumption (x_t), along with the consumptions of all FUTURE generations (x_{t+i}). Future generations’ utilities are discounted exponentially all multiplied by the constant β which measures “the degree to which the present generation values other peoples’ consumption relative to their own” (p.186). That is, $\beta = 1$ represents “perfect altruism” and $0 < \beta < 1$ “imperfect altruism” (ibid). Phelps and Pollak use this functional form in a dynamic game to characterize how more or less optimal saving rates depend on the degree of altruism. Laibson’s main interpretational move is to turn Phelps and Pollak’s generations into “temporal selves” (Laibson 1994, p.13), thereby proposing a *multiple-selves model*. The PRESENT self, denoted by t , evaluates a plan ($U^t(X)$) by its own undiscounted utility ($u(x_t)$), the exponentially discounted utilities of the FUTURE selves ($\sum_{i=1}^{\infty} \delta^i u(x_{t+i})$), and the anticipation that they will be biased (β) towards THEIR PRESENT if they re-evaluate the same plan. In short, the formal similarities in two of the three dimensions

allows Laibson to use a functional form originally designed for social preferences of generations to model time preferences of single decision makers.²³

In the same spirit, psychologist John Doyle (2013, p.125) remarks that quasi-hyperbolic discounting models “capture a psychological distinction that the other models cannot”, namely the “*qualitative* difference between now and any time to come, which we all appreciate intuitively, and which language mirrors in verb tenses”. Doyle further argues this phenomenon is more general as “[r]ewards that are *certain*, *mine* and *now*”, i.e., consequences in the Rastier’s identity zones of the three dimensions, “are all over-valued” (ibid).

These interdisciplinary sympathetic comments from a psychologist contrast with Ainslie’s (2012). According to him, sophisticated techniques of self-control are not captured by models where the whole of the preference reversal comes from the anticipation of a sudden “ β spike” of utility for a given consequence (ibid). Furthermore, because the shape of β - δ curves in the long-run are similar to the shape of exponential curves, they have just as much trouble to explain the great diversity in empirical measures of discount rates (see Frederick, Loewenstein and O’Donoghue 2002, table 1; Wilkinson and Klaes 2012, chaps.7-8). The general pattern is that discount rates decline over time, though there are many exceptions. Ainslie argues that β - δ curves reproduce this qualitative pattern but not quantitatively as well as hyperbolic discount curves and (unlike hyperbolic curves) without accounting for the exceptions. Finally, Ainslie’s own analysis of hyperbolic discounting draws on dynamic game theory (see esp. 1992) with a position on issues underlying multiple equilibria different from economists’. As he puts it, “[i]nformal discussion has suggested that the obstacle [in the adoption of hyperbolic models by economists] is economists’ determination to predict unique decisions, in principle at least, from a given set of prior motives” (2012, p.21). However, it may be that “a technically chaotic mechanism [...] is the way of the world” (2012, p.29). Indeed, at the beginning of his dissertation, Laibson states that “[t]he primary goal of this thesis is to formalize and extend Ainslie’s psychological analysis, by explicitly modeling individual decision-making as an intra-personal game” (1994, p.10). His interpretation of the multiple equilibria result is that “the model generates heterogeneous behavior without making recourse to heterogeneous preferences” (1994, p.40).

²³Elster (1979, p.71ff) should be acknowledged as a precedent for the intuition of adapting Phelps and Pollak’s model at the individual level. There are also precedents to Laibson’s introduction of Ainslie’s work in behavioral economics, especially Thaler (1981) from an empirical perspective and Thaler and Shefrin (1981) and Loewenstein and Prelec (1992) from a theoretical perspective.

Furthermore, two decision makers with the same preferences may end up in equilibria that are not Pareto indifferent, representing “bad habits” (when Pareto-inferior) or “self-control” (when Pareto-superior) (ibid). This reasoning is indeed close to Ainslie’s. However, not only does this interpretation disappear in the 1997 paper, but Laibson has put some efforts from the 2000s onwards with Christopher Harris to get a unique equilibrium with β - δ preferences (see Harris and Laibson 2013 for the first published version). In other words, multiple equilibria in hyperbolic and quasi-hyperbolic models trigger negative epistemic value judgments and positive ones from psychologists.

The asymmetry between gains and losses in sign effects attracts neither the theoretical attention nor the controversies documented above for risk preferences. Yet, in several joint and individual papers, Loewenstein and Prelec introduce hyperbolic (*not* quasi-hyperbolic) discounting in economics with a specific focus on sign effects. Their formal contribution is a generalized hyperbola that nests exponential, quasi-hyperbolic and hyperbolic discounting function as special cases (see esp. Loewenstein and Prelec 1992). They introduce hyperbolic discounting in economics by drawing on Kahneman and Tversky’s prospect theory to formalize the notion of temporal reference-dependence. This notably implies different objects of choice: *temporal prospects* as adjustment to consumption rather than consumption *per se*. They remark at some point that “interpersonal comparisons of material outcomes reflect a concern both for absolute and relative differences between one’s own and other persons’ payoffs” (Prelec and Loewenstein 1991, p.784). As it has not been developed further, Prelec and Loewenstein’s work remains a promising though neglected path to model the three dimensions altogether.

By contrast with risk preferences, there are some ambiguities between the positive/normative distinction at the individual level and at the level of subareas in economics (i.e., as what distinguishes ‘positive economics’ *versus* ‘standard economics’). In his landmark contribution on exponential discounting, Paul Samuelson’s (1937) original comments at the individual level are at best agnostic. He stresses the arbitrariness of the functional form throughout and only highlights the rationality underlying dynamic consistency for the theoretical case of perfect markets with interest rate. At the level of subareas in economics, he is more explicit on exponential discounting having no place in normative economics: “any connection between utility as discussed here and any welfare concept is disavowed [...] [and] any influence upon ethical judgments of

policy is one which deserves the impatience of modern economists” (p.161). Some have interpreted the first part of this quote as belonging to the individual level of the positive/normative issue (e.g., Frederick et al. 2002, p.355). This is not a fallacy *per se* because there are several methodological arguments to connect both levels (see Sen 1987; Hands 2012; Hausman and McPherson 2006). However, it goes against the general tendency shared by both standard and behavioral economists to derive value judgment of rationality from exponential discounting because it prevents intertemporal preference reversals (see, e.g., Gollier 2001; Loewenstein et al. 2015, pp.61-65). And symmetrically, value judgments of irrationality are derived from quasi-hyperbolic discounting because it implies such reversals. On this point, there are some subtleties in the use of quasi-hyperbolic discounting to model “sophisticated” *versus* “naive” decision makers. Recall that in Laibson’s models β captures the evaluation of a plan by a self at time t who *anticipates* all the next $t + 1$ selves to be biased towards THEIR PRESENT if re-evaluating the same plan. If these anticipations are assumed to be correct, then the multiple selves model is taken to be about a sophisticated decision maker; if incorrect, it is about a naive one.²⁴

Regarding other people

Turning to social preferences, two contrasts with risk and time preferences are worth noting at the outset. On the one hand, though some inspirations are drawn here and there from social psychology, there is no substantial influence from one psychologist (or even from a well-identified set of psychologists). On the other hand, there is not one but *several* – at least five – alternatives to the standard model (Rabin 1993; Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Andreoni and Miller 2002; Charness and Rabin 2002). Because of this diversity, an informal discussion of how non-self-centered motives are introduced in these models makes more sense than a discussion of their formalism (for which see Camerer 2003, chap.3; Sobel 2005, sect.3; Fehr and Schmidt 2006, sect.3; Wilkinson and Klaes 2012, sects.10.5-6).

Different ways of introducing non-self-centered motives in utility functions can be interpreted

²⁴In a seemingly unrelated contributions to the one with Phelps (though in the very same issue of the same journal and just after it), Pollak (1968) introduces the “sophisticate”/“naïve” distinction. In behavioral economics, O’Donoghue and Rabin (1999a; b; 2001; 2003; 2005; 2006) provide systematic contrasts between the sophisticated and the naïve cases. Much of Ainslie’s work focus on self-control and the normative medical implications of pathological self-control problems; the complex implications for standard and behavioral economics of the normal/pathological categories are not discussed here (see Vallois 2011; 2012a; b; 2014).

as specific cases of what Sen (1977) calls “sympathy”, i.e., different ways by which the decision maker’s utility is dependent on “the state of others” (2002, chap.1, p.35). For a given allocation, such dependence can be (1) on whether THE OTHER gets more or less than the decision maker (Fehr and Schmidt 1999), (2) on whether THE OTHERS get on average more or less than the decision maker (Bolton and Ockenfels 2000), (3) on THE OTHER(S) who get(s) the least, or on (4) how much WE get in total (both Andreoni and Miller 2002 and Charness and Rabin 2002 propose characterizations (2) and (3)). These four possibilities correspond to positions in ethics, moral, or political philosophy: (1) and (2) are two species of egalitarianism and (3)/(4) represents the Rawlsian/utilitarianist contrast. The decision maker’s utility can also depend on (5) THE OTHER’s (a) own motives and (b) beliefs about the decision maker’s motives (Rabin 1993). This possibility belongs to the realm of psychological game theory and does not match a position in ethics, moral, or political philosophy (see Dufwenberg 2008). In terms of Sen’s conceptual apparatus, (1)-(5) violate the requirement of self-welfare goal (i.e., that own welfare maximization is the unique goal). This is the case because they all imply that the decision maker is not *only* maximizing her or his own welfare, though they all imply that the decision maker is nevertheless *also* maximizing his or her own welfare. This explains why violations of self-centeredness in (1)-(5) are always accompanied with statements about the rationality of the underlying behaviors. Hence there is a sharp contrast with the other two dimensions, i.e., with violations of the independence axiom and dynamic consistency. More precisely, this *value* judgment of rationality comes from the theoretical *conventions* of using formally well-behaved utility functions, the maximization of which allows to derive game-theoretical equilibria to explain and predict *factual* behaviors in experimental games (see esp. Andreoni and Miller 2002; and Fehr and Schmidt 2006, sect.2.2).

All (1)-(5) insist that one contribution of their formal model is the possibility to capture decision makers’ value judgments: “the concerns for fairness and equity *of the economic actors being studied*” without “incorporat[ing] economists’ judgments of fairness and equity” (Rabin 1993, p.1282, my emphasis). In short, the use of formal language by the economist captures how decision makers would justify their behaviors in ordinary language. Because behavioral economists do not rely on verbal data, the methodological condition of possibility here is the sharing of ordinary language between economists and economic agents. Hence behavioral economists’ in-

terpretations of their models of social preferences in folk-psychological terms, of which *fairness*, *equity*, *altruism*, and *reciprocity* are the prominent ones in (1)-(5).

Is it possible for these models to go beyond sympathy and capture what Sen calls “commitment” in which there is no “tight link between individual welfare (with or without sympathy) and the choice of action” (2002, *ibid*)? A charitable answer is that, this is not the case but it was never meant to be the case. Though they put it differently, behavioral economists working on social preferences aim to capture “how people’s attitudes toward fairness influence their behavior and well-being” (Rabin 1993, *ibid*). This is empirically motivated by regularities from laboratory experiments with designs as neutral as possible, unlike the ones about social distance presented in section 1. More in line with Sen’s original criticisms, the impossibility to capture cases of commitment is problematic for the explanations of real-world phenomena extrapolated qualitatively from (1)-(5), which are present in the papers referenced above (see also Meier 2007). However, because (1)-(5) capture various degrees of Sen’s sympathy, they do partly respond to Sen’s original criticism by not picturing the decision maker as “close to being a social moron” anymore (Sen 1977, p.336). To go further and account for Sen’s commitment, “empirical evidence [...] cannot be sought in the mere observation of actual choices, and must involve other sources of information, including introspection and *discussion*” (1977, p.342, my emphasis). This is why Shiell and Rush (2003), in the only quasi-experimental study that looks for Sen’s commitment (to the best of my knowledge), need verbal data (collected through interviews with decision makers) to complement choice data so as to make commitment observable.²⁵

Behavioral economists often remark that including “specific economic variables like rules of the game, as well as social variables like the level of anonymity, the sex of one’s opponent, or the framing of the decision” (Andreoni and Miller 2002, p.738) is in the agenda for future theoretical developments. It is also not rare to read that the notion of identity (i.e, the decision makers’ personal, social and/or political identities) is central for general models of social preferences (e.g., Camerer 2003, p.476; see also Fehr and Schmidt 2006). Arguably, this project is conducted in the influential work of Roland Bénabou and Jean Tirole (2011) which extends their earlier contributions on non-monetary incentives and self-control to social preferences as understood in (1)-(5). Three broad aspects of Bénabou and Tirole’s work are worth commenting. Firstly,

²⁵On Sen’s commitment and the conceptual issues it raises, see the contributions in Peter and Schmid (2007).

by contrast with (1)-(5), Bénabou and Tirole do not emphasize that social preferences are necessarily rational (though they do not pronounce *explicit* value judgments of irrationality either, most of their remarks point in this direction). Secondly, as in (1)-(5), references from social psychology are only briefly mentioned and there is no focal psychologist(s) (in the sense in which Kahneman and Tversky and Ainslie are focal for risk and time preferences). Unlike in (1)-(5), however, Bénabou and Tirole (2011) mention much more references from social psychology, notably to show how their models allow to represent phenomena observed by social psychologists and to derive their economic implications. Finally, the third aspect is related to the central role of the notion of identity in these interpretations and economic implications. At one point, they argue that Sen's commitment, especially when explained by decision makers' identities, "can be formalized, in a way consistent with consequentialist rationality" through their model (Bénabou and Tirole 2011, fn32 pp.828-9). Hence the only way Sen's commitment can be integrated into the social preferences literature around behavioral economics is if it is devoided from one of its initial ambition, namely to capture some non-consequentialist form of rationality relevant for the behavioral foundations of economics.

Here are the conclusions one can draw from comparing behavioral economics' theoretical alternatives in the three dimensions altogether. First, consider the issue of interdisciplinarity. In models of risk and time preferences, central scientific terms borrowed from psychologists (e.g., loss aversion, impulsivity) are at the same time folk-psychological in themselves and close to the ones used by economists (e.g., risk aversion, impatience). In models of social preferences, there are no psychological terms borrowed from psychology; fairness, equity, altruism and reciprocity are all purely folk-psychological terms. Furthermore, these terms are at least ethically non-trivial if not purely ethical as with fairness. Second, consider the positive/normative issue. Standard models of risk and time preferences are normative benchmarks from which value judgments of irrationality are derived concerning both behavioral models and the empirical behaviors they are about. By contrast, standard models of social preferences are not normative benchmarks; behavioral models of social preferences are considered to be about rational behaviors. Indeed, as we have seen, behavioral economists seem to wish to avoid calling irrational a behavior that the decision maker would describe as ethical.

Conclusion

This paper tried to provide a broad picture of the methodological issues underlying how behavioral economics is changing three central dimensions of economic rationality – risk, time and other people. A new perspective on these issues has been proposed by following the entanglement thesis of Putnam, Walsh and Sen, for whom facts, values and conventions have interdependent meanings that can be understood by scrutinizing formal and ordinary language uses. Rastier’s table helped to systematize the picture by looking at linguistic distinctions that mark three zones within each dimension. Empirical and theoretical contributions within each dimension involve interactions between different zones, but also the implicit presence of at least one zone from each of the other two dimensions, i.e., FOR SURE, NOW or ME. A common feature of these contributions is to depart from the consequentialism inherent to all standard models, a great source of value judgments in economics. Handling these departures theoretically involves the use of more or less folk-psychological terms that are never purely scientific or technical from psychology.

The broad picture of behavioral economics presented here may be relevant for further work on three topics of historical, methodological and philosophical interests. The first one is the revival of a type of experiments discussed by Guala (2000) and that Mongin (2014) calls “the experimental translation of rationality” (see also Mongin 2017; Eli 2017). Roughly, it consists in an early set of experiments on behavior under risk designed to test the normative status of expected utility theory and eventually to modify it in the light of empirical evidence from individuals violating the models and then reflecting on their own choices. Notice that, on the one hand, there is an explicit and controlled violation of the fact/value dichotomy, and, on the other hand, it seems extendable to decisions over time and regarding other people. The second topic is the emergence of reflections on the role of narratives in individual behavior, i.e., the role of the stories we tell and are told about the world and ourselves in shaping our behavior (see Davis 2009; 2011; Ross 2005; 2014; Akerlof and Snower 2016; Falk and Tirole 2016; Shiller 2017). Conceptually, narratives can be characterized as descriptions of events that are possibly meant to influence the behavior of an individual by changing his or her values. However, in this literature it is not very clear what empirically counts as one narrative or what elements of

narratives should be of empirical significance. One suggestion from the broad picture of this paper would be to look at how the three dimensions – and their zones – are articulated in a narrative. Finally, the third topic has already been stated in the introduction: the current shift towards contributions across the three dimensions. It can be remarked that, in the contributions within the three dimensions, there are more similarities between the dimensions of risk and time than between the dimensions of other people and either one of the other two. This carries over in contributions across the three dimensions in the sense that the risk-time pair is by far the one that receives the most attention. There also, the entanglement thesis provides a fruitful perspective to bring together the social dimension of standard models with underlying methodological issues common to the other two dimensions.

References

- Abdellaoui, M., Barrios, C. and Wakker, P. P.: 2007, Reconciling Introspective Utility With Revealed Preference: Experimental Arguments Based on Prospect Theory, *Journal of Econometrics* **138**(1), 356–378.
- Abdellaoui, M., Bleichrodt, H., L'Haridon, O. and Paraschiv, C.: 2013, Is There One Unifying Concept of Utility? An Experimental Comparison of Utility Under Risk and Utility Over Time, *Management Science* **59**(9), 2153–2169.
- Ainslie, G.: 1975, Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control, *Psychological Bulletin* **82**(4), 463–496.
- Ainslie, G.: 1992, *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*, Cambridge University Press, Cambridge, England.
- Ainslie, G.: 2012, Pure Hyperbolic Discount Curves Predict "eyes open" Self-control, *Theory and Decision* **73**(1), 3–34.
- Akerlof, G. A. and Snower, D. J.: 2016, Bread and Bullets, *Journal of Economic Behavior & Organization* **126**(Part B), 58–71.

- Andersen, S., Harrison, G. W., Lau, M. I. and Rutström, E. E.: 2008, Eliciting Risk and Time Preferences, *Econometrica* **76**(3), 583–618.
- Andreoni, J. and Miller, J.: 2002, Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism, *Econometrica* **70**(2), 737–753.
- Andreoni, J. and Sprenger, C.: 2012, Risk Preferences Are Not Time Preferences, *American Economic Review* **102**(7), 3357–3376.
- Angner, E.: 2015, "To Navigate Safely in the Vast Sea of Empirical Facts". Ontology and Methodology in Behavioral Economics, *Synthese* **192**(11), 3557–3575.
- Angner, E. and Loewenstein, G.: 2012, Behavioral Economics, in U. Mäki (ed.), *Handbook of the Philosophy of Science, Vol.5: Philosophy of Economics*, Elsevier, Oxford, pp. 641–690.
- Antinyan, A.: 2014, Loss and Other-Regarding Preferences, *Working Paper* .
URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2413022
- Atkinson, G., Dietz, S., Helgeson, J., Hepburn, C. and Sælen, H.: 2009, Siblings, Not Triplets: Social Preferences for Risk, Inequality and Time in Discounting Climate Change, *Economics: The Open-Access, Open-Assessment E-Journal* **3**, 0–29.
URL: <http://www.economics-ejournal.org/economics/journalarticles/2009-26>
- Aumann, R. J. and Dreze, J. H.: 2009, Assessing Strategic Risk, *American Economic Journal: Microeconomics* **1**(1), 1–16.
- Barberis, N., Huang, M. and Thaler, R. H.: 2006, Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing, *The American Economic Review* **96**(4), 1069–1090.
- Bardsley, N.: 2008, Dictator Game Giving: altruism or artefact?, *Experimental Economics* **11**(2), 122–133.
- Bardsley, N., Cubitt, R. P., Loomes, G., Moffatt, P., Starmer, C. and Sugden, R.: 2010, *Experimental Economics: Rethinking the Rules*, Princeton University Press, Princeton.
- Baron, J.: 1994, Nonconsequentialist Decisions, *Behavioral and Brain Sciences* **17**(1), 1–42.

- Baucells, M. and Heukamp, F. H.: 2012, Probability and Time Trade-off, *Management Science* **58**(4), 831–842.
- Bénabou, R. and Tirole, J.: 2011, Identity, Morals, and Taboos: Beliefs as Assets, *The Quarterly Journal of Economics* **126**(2), 805–855.
- Blaug, M.: 1998, The Positive-normative Distinction, in J. B. Davis, W. Hands and U. Mäki (eds), *Handbook of Economic Methodology*, Elgar, Cheltenham, pp. 370–4.
- Bleichrodt, H., Rohde, K. I. and Wakker, P. P.: 2008, Koopman’s Constant Discounting for Intertemporal Choice: A Simplification and a Generalization, *Journal of Mathematical Psychology* **52**(6), 341–347.
- Bohnet, I. and Frey, B. S.: 1999, Social Distance and Other-regarding Behavior in Dictator Games: Comment, *The American Economic Review* **89**(1), 335–339.
- Bolton, G. E. and Ockenfels, A.: 2000, ERC: A Theory of Equity, Reciprocity, and Competition, *The American Economic Review* **90**(1), 166–193.
- Bolton, G. E. and Ockenfels, A.: 2010, Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment, *The American Economic Review* **100**(1), 628–633.
- Brañas-Garza, P.: 2006, Poverty in Dictator Games: Awakening Solidarity, *Journal of Economic Behavior & Organization* **60**(3), 306–320.
- Broome, J.: 1991, *Weighing Goods: Equality, Uncertainty and Time*, Basil Blackwell, Oxford.
- Bughart, D., Epper, T. and Fehr, E.: 2015, The Two Faces of Independence: Betweenness and Homotheticity, *Working Paper*.
- URL:** http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2523486
- Burks, A. W.: 1977, *Chance, Cause, Reason. An Inquiry into the Nature of Scientific Evidence*, The University of Chicago Press, Chicago and London.
- Camerer, C. F.: 2003, *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, Princeton.

- Camerer, C. F., Loewenstein, G. F. and Rabin, M.: 2004, *Advances in Behavioral Economics*, Princeton University Press, Princeton.
- Castro Caldas, J. and Neves, V.: 2012, Introduction. The Issues at Stake, in J. Castro Caldas and V. Neves (eds), *Facts, Values and Objectivity in Economics*, Routledge, London and New York.
- Charness, G. and Rabin, M.: 2002, Understanding Social Preferences with Simple Tests, *The Quarterly Journal of Economics* **117**(3), 817–869.
- Cozic, M.: 2009, Du concept de conséquence logique: présentation, in D. Bonnay and M. Cozic (eds), *Textes clés de philosophie de la logique*, Vrin, Paris, pp. 75–82.
- Danziger, K.: 1997, *Naming the Mind: How Psychology Found its Language*, Sage Publications, London.
- Davis, J. B.: 2009, Identity and Individual Economic Agents: A Narrative Approach, *Review of Social Economy* **67**(1), 71–94.
- Davis, J. B.: 2011, *Individuals and identity in economics*, Cambridge University Press, New York.
- Deaton, A. and Muellbauer, J.: 1980, *Economics and Consumer Behavior*, Cambridge University Press, New York.
- DellaVigna, S.: 2009, Psychology and Economics: Evidence from the Field, *Journal of Economic Literature* **47**(2), 315–372.
- Diecidue, E. and Wakker, P. P.: 2001, On the Intuition of Rank-Dependent Utility, *Journal of Risk and Uncertainty* **23**(3), 281–298.
- Doyle, J. R.: 2013, Survey of Time Preference, Delay Discounting Models, *Judgment and Decision Making* **8**(2), 116–135.
- Dufwenberg, M.: 2008, Psychological Games, in S. N. Durlauf and L. E. Blume (eds), *The New Palgrave Dictionary of Economics. Second Edition. Volume 6.*, Palgrave, Hampshire, pp. 714–717.

- Egidi, M.: 2012, The Cognitive Explanation of Economic Behaviour: From Simon to Kahneman, in R. Arena, A. Festré and N. Lazaric (eds), *The Handbook of Economics and Knowledge*, Edward Elgar Publishing, Cheltenham, pp. 183–210.
- Eli, V.: 2017, Testing and Justifying Normative Theories of Decision Making: An experimental approach to Rationality, *PhD Dissertation (in preparation at HEC, Paris)*.
- Ellingsen, T.: 1994, Cardinal Utility: A History of Hedonimetry, in M. Allais and O. Hagen (eds), *Cardinalism*, Springer, Dordrecht, pp. 105–166.
- Ellsberg, D.: 1954, Classic and Current Notions of "Measurable Utility", *Economic Journal* **64**, 528–556.
- Elster, J.: 1979, *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge University Press, Cambridge, England.
- Engel, C.: 2011, Dictator Games: A Meta Study, *Experimental Economics* **14**(4), 583–610.
- Epper, T. and Fehr-Duda, H.: 2015, The Missing Link: Unifying Risk Taking and Time Discounting, *Working Paper*.
- URL:** http://thomasepper.com/papers/wp/missing_link.pdf
- Falk, A. and Tirole, J.: 2016, Narratives, Imperatives and Moral Reasoning, *Working Paper*.
- URL:** https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/by/tirole/falk_tirole_moral_reasoning_18051
- Faralla, V., Benuzzi, F., Nichelli, P. and Dimitri, N.: 2012, Gains and Losses in Intertemporal Preferences: A Behavioral Study, in A. Innocenti and A. Sirigu (eds), *Neuroscience and the Economics of Decision Making*, Routledge, London and New York.
- Fehr, E. and Schmidt, K. M.: 1999, A Theory of Fairness, Competition, and Cooperation, *The Quarterly Journal of Economics* **114**(3), 817–868.
- Fehr, E. and Schmidt, K. M.: 2006, The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories, in S.-C. Kolm and J.-M. Ythier (eds), *Handbook of the Economics of Giving, Altruism and Reciprocity: Foundations (Vol.1)*, Elsevier, Amsterdam, pp. 615–691.

- Fennema, H. and Wakker, P. P.: 1997, Original and Cumulative Prospect Theory: A Discussion of Empirical Differences, *Journal of Behavioral Decision Making* **10**(1), 53–64.
- Fishburn, P. C.: 1989, Retrospective on the Utility Theory of von Neumann and Morgenstern, *Journal of Risk and Uncertainty* **2**(2), 127–158.
- Fishburn, P. C. and Rubinstein, A.: 1982, Time Preference, *International Review of Economics* **23**(3), 677–694.
- Fishburn, P. C. and Wakker, P. P.: 1995, The Invention of the Independence Condition for Preferences, *Management Science* **41**(7), 1130–1144.
- Fontaine, P.: 2000, Making Use of the Past: Theorists and Historians on the Economics of Altruism, *The European Journal of the History of Economic Thought* **7**(3), 407–422.
- Frederick, S., Loewenstein, G. and O’Donoghue, T.: 2002, Time Discounting and Time Preference: A Critical Review, *Journal of Economic Literature* **40**(2), 351–401.
- Fudenberg, D.: 2006, Advancing Beyond Advances in Behavioral Economics, *Journal of Economic Literature* **XLIV**(September), 694–711.
- Gabbay, D. M. and Woods, J.: 2005, The Practical Turn in Logic, in D. M. Gabbay and F. Guenther (eds), *Handbook of Philosophical Logic, Volume 13*, 2 edn, Springer, pp. 15–122.
- Gajdos, T. and Weymark, J. A.: 2012, Introduction to Inequality and Risk, *Journal of Economic Theory* **147**(4), 1313–1330.
- Galison, P.: 1999, Trading Zone: Coordinating Action and Belief, in M. Biagioli (ed.), *The Science Studies Reader*, Routledge, New York and London, pp. 137–160.
- Geiger, N.: 2016, The Rise of Behavioural Economics: A Quantitative Assessment, *Working Paper* .
URL: <https://ideas.repec.org/p/zbw/hohpro/442015.html>
- Gollier, C.: 2001, *The Economics of Risk and Time*, The MIT Press, Cambridge, MA and London.

- Gorman, W. M.: 1968, The Structure of Utility Functions, *The Review of Economic Studies* **35**(4), 367–390.
- Grüne-Yanoff, T.: 2015, Models of Temporal Discounting 1937-2000: An Interdisciplinary Exchange between Economics and Psychology, *Science in Context* **28**(4), 675–713.
- Guala, F.: 2000, The Logic of Normative Falsification: Rationality and Experiments in Decision Theory, *Journal of Economic Methodology* **7**(1), 59–93.
- Güth, W., Levati, V. and Ploner, M.: 2008, On the Social Dimension of Time and Risk Preferences: An Experimental Study, *Economic Inquiry* **46**(2), 261–272.
- Halevy, Y.: 2008, Strotz Meets Allais: Diminishing Impatience and the Certainty Effect, *American Economic Review* **98**(3), 1145–1162.
- Hammond, P. J.: 1976, Changing Tastes and Coherent Dynamic Choice, *Review of Economic Studies* **43**(1), 159–173.
- Hammond, P. J.: 1977, Dynamic Restrictions on Metastatic Choice, *Economica* **44**(176), 337–350.
- Hammond, P. J.: 1983, Ex-post Optimality as a Dynamically Consistent Objective for Collective Choice Under Uncertainty, in P. K. Pattanaik and M. Salles (eds), *Social Choice and Welfare*, North-Holland, Amsterdam, New York and Oxford, pp. 175–205.
- Hammond, P. J.: 1987, Altruism, in J. M. Eatwell and P. Newman (eds), *The New Palgrave Dictionary of Economics*, Palgrave, London, pp. 85–87.
- Hammond, P. J.: 1988a, Consequentialism and the Independence Axiom, in B. Munier (ed.), *Risk, Decision and Rationality*, Reidel, Dordrecht, pp. 503–516.
- Hammond, P. J.: 1988b, Consequentialist Foundations for Expected Utility Theory, *Theory and Decision* **25**, 25–78.
- Hammond, P. J.: 1989, Consistent Plans, Consequentialism, and Expected Utility, *Econometrica* **57**(6), 1445–1449.

- Hammond, P. J.: 1998, Objective expected utility, *in* S. Barberà, P. J. Hammond and C. Seidl (eds), *Handbook of utility Theory*, Kluwer, Dordrecht, chapter 5.
- Hammond, P. J. and Zank, H.: 2014, Rationality and Dynamic Consistency Under Risk and Uncertainty, *in* M. J. Machina and W. K. Viscusi (eds), *Handbook of the Economics of Risk and Uncertainty*, Elsevier, Amsterdam, pp. 41–98.
- Hands, W. D.: 2012, The Positive-Normative Dichotomy and Economics, *in* U. Mäki (ed.), *Handbook of the Philosophy of Science, Vol.5: Philosophy of Economics*, Elsevier, Oxford, pp. 219–239.
- Hands, W. D.: 2015, Normative rational choice theory: Past, present, and future, *Working Paper*.
- URL:** https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1738671
- Harris, C. and Laibson, D.: 2013, Instantaneous Gratification, *The Quarterly Journal of Economics* **128**(1), 205–248.
- Harrison, G. W., Lau, M. I. and Rutström, E. E.: 2007, Estimating Risk Attitudes in Denmark: A Field Experiment, *The Scandinavian Journal of Economics* **109**(2), 341–368.
- Harsanyi, J. C.: 1955, Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility, *The Journal of Political Economy* **63**(4), 309–321.
- Harsanyi, J. C.: 1988, Assessing Other People’s Utilities, *in* B. R. Munier (ed.), *Risk, Decision and Rationality*, Reidel, Dordrecht, pp. 127–138.
- Hausman, D. M. and Mcpherson, M. S.: 2006, *Economic Analysis, Moral Philosophy, and Public Policy (Second Edition)*, Cambridge University Press, Cambridge, England.
- Heidl, S.: 2016, *Philosophical Problems of Behavioural Economics*, Routledge, London and New York.
- Heukelom, F.: 2009, *Kahneman and Tversky and the Making of Behavioral Economics (phD thesis)*, PhD thesis, Amsterdam.

- Heukelom, F.: 2011, Building and Defining Behavioral Economics, *Research in the History of Economic Thought and Methodology* **29-A**, 1–29.
- Heukelom, F.: 2012, Behavioral economics, in J. B. Davis and W. D. Hands (eds), *The Elgar Companion to Recent Economic Methodology*, Edward Elgar Publishing, chapter 2, pp. 19–38.
- Heukelom, F.: 2014, *Behavioral Economics: A history*, Oxford University Press, New York.
- Hoffman, E., McCabe, K. and Smith, V. L.: 1996, Social Distance and Other-regarding Behavior in Dictator Games, *The American Economic Review* **86**(3), 653–660.
- Ida, T. and Ogawa, K.: 2013, Inequality Aversion, Social Discount, and Time Discount Rates, *Working Paper* .
URL: <https://www.econ.kyoto-u.ac.jp/projectcenter/Paper/e-10-013.pdf>
- Jolls, C., Sunstein, C. R. and Thaler, R. H.: 1998, A Behavioral Approach to Law and Economics, *Stanford Law review* (50), 1471–1550.
- Jones, B. and Rachlin, H.: 2006, Social Discounting, *Psychological Science* **17**(4), 283–286.
- Kahneman, D., Knetsch, J. L. and Thaler, R. H.: 1986, Fairness and the Assumptions of Economics, *The Journal of Business* **59**(4), S285–S300.
- Kahneman, D. and Tversky, A.: 1979, Prospect Theory: An Analysis of Decision under Risk, *Econometrica* **47**(2), 263–291.
- Kao, Y.-F. and Velupillai, V.: 2015, Behavioural economics: classical and modern, *The European Journal of the History of Economic Thought* **22**(223-271).
- Klaes, M. and Sent, E.-M.: 2005, A Conceptual History of the Emergence of Bounded Rationality, *History of Political Economy* **37**(1), 27–59.
- Köszegi, B. and Rabin, M.: 2006, A Model of Reference-Dependent Preferences, *The Quarterly Journal of Economics* **121**(4), 1133–1165.
- Köszegi, B. and Rabin, M.: 2007, Reference-Dependent Risk Attitudes, *The American Economic Review* **97**(4), 1047–1073.

- Kőszegi, B. and Rabin, M.: 2009, Reference-Dependent Consumption Plans, *American Economic Review* **99**(3), 909–936.
- Kovarik, J.: 2009, Giving it Now or Later: Altruism and Discounting, *Economics Letters* **102**(3), 152–154.
- Kreps, D. and Porteus, E.: 1978, Temporal Resolution of Uncertainty and Dynamic Choice Theory, *Econometrica* **46**(1), 185–200.
- Laibson, D.: 1994, Hyperbolic Discounting and Consumption, (Ph.D dissertation, MIT).
- Laibson, D.: 1997, Golden Eggs and Hyperbolic Discounting, *The Quarterly Journal of Economics* **112**(2), 443–478.
- Lapied, A. and Renault, O.: 2012, An Investigation of Time Consistency for Subjective Discounted Utility, *Working Paper* .
URL: <https://halshs.archives-ouvertes.fr/halshs-00793174/>
- Leliveld, M. C., van Beest, I., van Dijk, E. and Tenbrunsel, A. E.: 2009, Understanding the Influence of Outcome Valence in Bargaining: A Study on Fairness Accessibility, Norms, and Behavior, *Journal of Experimental Social Psychology* **45**(3), 505–514.
- Levine, D. K.: 2012, *Is Behavioral Economics Doomed?: The Ordinary versus the Extraordinary*, Open Book Publishers, Cambridge, England.
URL: <http://www.openbookpublishers.com/product/77>
- Levitt, S. D. and List, J. A.: 2007, What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World ?, *Journal of Economic Perspectives* **21**(2), 153–174.
- List, J. A.: 2007, On the Interpretation of Giving in Dictator Games, *Journal of Political Economy* **115**(3), 482–493.
- Loewenstein, G., O'Donoghue, T. and Bhatia, S.: 2015, Modeling the Interplay Between Affect and Deliberation, *Decision* **2**(2), 55–81.
- Loewenstein, G. and Prelec, D.: 1992, Anomalies in Intertemporal Choice: Evidence and an Interpretation, *The Quarterly Journal of Economics* **107**(2), 573–597.

- MacCrimmon, K. R. and Larson, S.: 1979, Utility Theory: Axioms vs Paradoxes, *in* M. Allais and O. Hagen (eds), *Expected Utility Hypothesis and the Allais Paradox*, Reidel, Dordrecht.
- Machina, M. J.: 1989, Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty, *Journal of Economic Literature* **XXVII**(December), 1622–1668.
- Marchetti, G. and Marchetti, S.: 2017, Behind and Beyond the Fact/Value Dichotomy, *in* G. Marchetti and S. Marchetti (eds), *Facts and Values: The Ethics and Metaphysics of Normativity*, Routledge, London and New York.
- Markowitz, H.: 1952, The Utility of Wealth, *The Journal of Political Economy* **IX**(2), 151–158.
- Mas-Colell, A., Whinston, M. D. and Green, J. R.: 1995, *Microeconomic Theory*, Oxford University Press, New York and Oxford.
- Meier, S.: 2007, A Survey of Economic Theories and Field Evidence on Pro-Social Behavior, *in* B. S. Frey and A. Stutzer (eds), *Economics and Psychology: A Promising New Cross-Disciplinary Field*, The MIT Press, Cambridge, MA and London, pp. 51–88.
- Mongin, P.: 2006a, A Concept of Progress for Normative Economics, *Economics and Philosophy* **22**(01), 19–54.
- Mongin, P.: 2006b, Value Judgments and Value Neutrality in Economics, *Economica* **73**(290), 257–286.
- Mongin, P.: 2009, Duhemian Themes in Expected Utility Theory, *in* A. Brenner and J. Gayon (eds), *French Studies in the Philosophy of Science*, Springer Netherlands, Dordrecht, pp. 303–357.
- Mongin, P.: 2014, Le Paradoxe d’Allais: Comment lui rendre sa signification perdue? (Allais’s Paradox: How to Give It Back Its Lost Meaning?), *Revue Economique* **65**(5), 743–779.
- Mongin, P.: 2017, Bref addendum sur la "conception expérimentale de la rationalité", *Revue Economique* **Forthcoming**.
- Moscatti, I.: 2013a, How Cardinal Utility Entered Economic Analysis, 1909-1944, *European Journal of the History of Economic Thought* **20**(6), 906–939.

- Moscatti, I.: 2013b, Were Jevons, Menger and Walras Really Cardinalists? On the Notion of Measurement in Utility Theory, Psychology, Mathematics and Other Disciplines, 1870-1910, *History of Political Economy* **45**(3), 373–414.
- Moscatti, I.: 2016a, How Economists Came to Accept Expected Utility Theory: The Case of Samuelson and Savage, *Journal of Economic Perspectives* **30**(Forthcoming).
- Moscatti, I.: 2016b, *Measuring Utility: from the Marginal Revolution to Neuroeconomics*, forthcoming edn, Oxford University Press, Oxford.
- Mullainathan, S. and Thaler, R. H.: 2001, Behavioral Economics, in N. J. Smelser and P. B. Baltes (eds), *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier Science Ltd., Amsterdam, pp. 1094–1100.
- Novakova, J. and Flegr, J.: 2013, How Much Is Our Fairness Worth? The Effect of Raising Stakes on Offers by Proposers and Minimum Acceptable Offers in Dictator and Ultimatum Games, *PLoS ONE* **8**(4), 1–9.
- O'Donoghue, T. and Rabin, M.: 1999a, Doing It Now or Later, *The American Economic Review* **89**(1), 103–124.
- O'Donoghue, T. and Rabin, M.: 1999b, Incentives for Procrastinators, *The Quarterly Journal of Economics* **114**(3), 769–816.
- O'Donoghue, T. and Rabin, M.: 2001, Choice and Procrastination, *The Quarterly Journal of Economics* **116**(1), 121–160.
- O'Donoghue, T. and Rabin, M.: 2003, Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes, *The American Economic Review* **93**(2), 186–191.
- O'Donoghue, T. and Rabin, M.: 2005, Optimal Taxes for Sin Goods, *Swedish Economic Policy* **12**(2), 7–39.
- O'Donoghue, T. and Rabin, M.: 2006, Optimal sin taxes, *Journal of Public Economics* **90**(10-11), 1825–1849.

- Pesendorfer, W.: 2006, Behavioral Economics Comes of Age: A Review Essay on Advances in Behavioral Economics, *Journal of Economic Literature* **XLIV**(September), 712–721.
- Peter, F. and Schmid, H. B.: 2007, *Rationality and Commitment*, Oxford University Press, Oxford.
- Phelps, E. S. and Pollak, R. A.: 1968, On Second-Best National Saving and Game-Equilibrium Growth, *The Review of Economic Studies* **35**(2), 185–199.
- Pollak, R. A.: 1968, Consistent Planning, *The Review of Economic Studies* **35**(2), 201–208.
- Pratt, J. W.: 1964, Risk Aversion in the Small and in the Large, *Econometrica* **32**(1), 122–136.
- Prelec, D. and Loewenstein, G.: 1991, Decision Making Over Time and Under Uncertainty: A Common Approach, *Management Science* **37**(7), 770–786.
- Putnam, H.: 2002, *The Collapse of the Fact/Value Dichotomy and Other Essays*, Harvard University Press, Cambridge, MA and London.
- Putnam, H.: 2004, *Ethics Without Ontology*, Harvard University Press, Cambridge, MA and London.
- Putnam, H.: 2013, *Philosophy in an Age of Science*, Harvard University Press, Cambridge, England.
- Putnam, H.: 2015, Naturalism, realism, and Normativity, *Journal of the American Philosophical Association* **1**(2), 312–328.
- Putnam, H. and Walsh, V.: 2011, *The End of Value-Free Economics*, Routledge, Abingdon.
- Quiggin, J.: 1982, A Theory of Anticipated Utility, *Journal of Economic Behavior & Organization* **3**(4), 323–343.
- Quiggin, J. and Wakker, P. P.: 1994, The Axiomatic Basis of Anticipated Utility, *Journal of Economic Theory* **64**(2), 486–499.
- Rabin, M.: 1993, Incorporating Fairness into Game Theory and Economics, *The American Economic Review* **83**(5), 1281–1302.

- Rabin, M.: 2000, Risk Aversion and Expected-Utility Theory: A Calibration Theorem, *Econometrica* **68**(5), 1281–1292.
- Rabin, M. and Thaler, R. H.: 2001, Anomalies: Risk Aversion, *Journal of Economic Perspectives* **15**(1), 219–232.
- Rastier, F.: 1996, Représentation ou interprétation?, in V. Rialle and D. Fissette (eds), *Penser l'esprit: des sciences de la cognition à une philosophie cognitive*, Presse Universitaire de Grenoble, Grenoble, pp. 219–232.
- Rastier, F.: 2008, Le langage sans origine ou l'émergence du milieu sémiotique, in R. Delamotte-Legrand, C. Hudelot and A. Salaza Orvig (eds), *Dialogues, mouvements discursifs, significations: hommage à Frédéric François*, Cortil-Wodon: E.M.E., Fernelmont, pp. 207–222.
- Rastier, F.: 2012, Langage et pensée : dualité sémiotique ou dualisme cognitif ?, *Texte !* **XVII**(1-2), 1–42.
- Reiss, J.: 2017, Fact-Value Entanglement in Positive Economics, *Journal of Economic Methodology* **24**(2), 134–149.
- Ross, D.: 2005, *Economic Theory and Cognitive Science: Microexplanation*, The MIT Press, Cambridge, MA and London.
- Ross, D.: 2014, *Philosophy of Economics*, Palgrave Macmillan, Basingstoke.
- Saito, K.: 2013, Social Preferences Under Risk: Equality of Opportunity versus Equality of Outcome, *American Economic Review* **103**(7), 3084–3101.
- Samuelson, P. A.: 1937, A Note on Measurement of Utility, *The Review of Economic Studies* **4**(2), 155–161.
- Schmidt, U.: 2003, Reference Dependence in Cumulative Prospect Theory, *Journal of Mathematical Psychology* **47**(2), 122–131.
- Schmidt, U., Starmer, C. and Sugden, R.: 2008, Third-generation Prospect Theory, *Journal of Risk and Uncertainty* **36**(3), 203–223.

- Sen, A. K.: 1977, Rational Fools: A Critique of the Behavioral Foundations of Economic Theory, *Philosophy and Public Affairs* **6**(4), 317–344.
- Sen, A. K.: 1987, *On Ethics & Economics*, Blackwell, Malden.
- Sen, A. K.: 2002, *Rationality and Freedom*, Harvard University Press, Cambridge, MA.
- Sen, A. K.: 2005, Walsh on Sen after Putnam, *Review of Political Economy* **17**(1), 107–113.
- Sen, A. K.: 2009, *The Idea of Justice*, Harvard University Press, Cambridge, MA.
- Sent, E.-M.: 2004, Behavioral Economics: How Psychology Made Its (Limited) Way Back Into Economics, *History of Political Economy* **36**(4), 735–760.
- Shiell, A. and Rush, B.: 2003, Can Willingness to Pay Capture the Value of Altruism? An Exploration of Sen’s Notion of Commitment, *Journal of Socio-Economics* **32**(6), 647–660.
- Shiller, R. J.: 2017, Narrative Economics, *The American Economic Review* **107**(4), 967–1004.
- Sobel, J.: 2005, Interdependent Preferences and Reciprocity, *Journal of Economic Literature* **XLIII**(2), 392–436.
- Stommel, E.: 2013, *Reference-Dependent Preferences: A Theoretical and Experimental Investigation of Individual Reference-Point Formation*, Springer.
- Strotz, R. H.: 1956, Myopia and Inconsistency in Dynamic Utility Maximization, *The Review of Economic Studies* **23**(3), 165–180.
- Tarski, A.: 2009, Du concept de conséquence logique, in D. Bonnay and M. Cozic (eds), *Textes clés de philosophie de la logique*, Vrin, Paris, pp. 83–98.
- Thaler, R. H.: 1980, Toward a Positive Theory of Consumer Choice, *Journal of Economic Behavior & Organization* **1**(1), 7–59.
- Thaler, R. H.: 1981, Some Empirical Evidence on Dynamic Inconsistencies, *Economics Letters* **8**(3), 201–207.
- Thaler, R. H.: 2015, *Misbehaving: the Making of Behavioral Economics*, W. W. Norton & Company, New York.

- Thaler, R. H. and Shefrin, H. M.: 1981, An Economic Theory of Self-Control, *The Journal of Political Economy* **89**(2), 392–406.
- Trautmann, S. T. and Vieider, F. M.: 2012, Social Influences on Risk Attitudes: Applications in Economics, in S. Roeser, R. Hillerbrand, P. Sandin and M. Peterson (eds), *Handbook of Risk Theory: Epistemological, Decision Theory, Ethics, and Social Implications of Risk*, Springer, Dordrecht, pp. 575–600.
- Tversky, A. and Kahneman, D.: 1992, Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty* **5**(4), 297–323.
- Vallois, N.: 2011, The Pathological Paradigm of Neuroeconomics, *Oeconomia* **1**(4), 525–556.
- Vallois, N.: 2012a, Des décideurs incohérents mais rationnels: la réhabilitation normative de l'incohérence séquentielle par la psychologie évolutionniste, *Working Paper* .
URL: <https://paris1.academia.edu/NicolasVallois>
- Vallois, N.: 2012b, *L'émergence d'un nouveau domaine de savoir: la neuroéconomie*, Ph.d, Paris 1 Panthéon-Sorbonne.
- Vallois, N.: 2014, Neurosciences et politiques publiques : Vers un nouvel interventionnisme économique, *Revue de Philosophie Economique* **15**(2), 132–175.
- Wakker, P. P.: 1994, Separating Marginal Utility and Probabilistic Risk Aversion, *Theory and Decision* **36**(1), 1–44.
- Wakker, P. P.: 2010, *Prospect Theory for Risk and Ambiguity*, Cambridge University Press, Cambridge, England.
- Wakker, P. P. and Tversky, A.: 1993, An Axiomatization of Cumulative Prospect Theory, *Journal of Risk and Uncertainty* **7**(2), 147–175.
- Walsh, V.: 1996, Rationality, Allocation, and Reproduction, *Technical report*, Oxford.
- Wilkinson, N. and Klaes, M.: 2012, *An Introduction to Behavioral Economics. 2nd Edition*, Palgrave Macmillan, Basingstoke.

Zeckhauser, R.: 2014, Preface 2, *in* M. J. Machina and W. K. Viscusi (eds), *Handbook of the Economics of Risk and Uncertainty*, Elsevier and North-Holland, Amsterdam, pp. xvii–xxix.