



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

Dorian Jullien

► To cite this version:

Dorian Jullien. Under Risk, Over Time and Regarding Other People: Rationality Across Three Dimensions. 2019. hal-03233897

HAL Id: hal-03233897

<https://hal.science/hal-03233897>

Preprint submitted on 25 May 2021

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 and Regarding Other People: Rationality Across Three Dimensions

Dorian Jullien*

January 25, 2019

Abstract

The paper contrasts the classical contributions in behavioral economics which investigated risk, time or social preferences independently of each others before the middle of the 2000s with the more recent contributions that jointly investigate these type of preferences by pair (e.g., how risk and time preference interact). It suggests that, while standard models could be used as normative benchmarks to judge some behavioral phenomena as non-rational in the classical contributions, this is more complicated in the recent contributions because interactions between types of preferences tend to decrease the normative force of standard models.

Keywords: behavioral economics, rationality, decision theory, risk preferences, time preferences, social preferences

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

*dorian.jullien@gredeg.cnrs.fr. Université Côte d’Azur, GREDEG (UMR 7321; 250 rue Albert Einstein - Batiment 2, 06560 Valbonne). Most passages of this paper are taken from the second chapter of my PhD dissertation and some of them (in the introduction and in the first section) have also appeared in a paper published in 2018 in *Research in History of Economic Thought and Method* (36:C, 121-158) entitled “Under Risk, Over Time and Regarding Other People: Language and Rationality Within Three Dimensions”.

“[A] rich toolbox for
 “neoclassical repair,” [...] may
 be a curse (a Pandora’s box)
 rather than a blessing [...].
 Some guidance on how risk
 attitudes, time preferences, and
 other-regarding concerns are
 interrelated becomes, therefore,
 necessary when we want to
 make sound behavioral
 predictions. [...] Except for a
 few attempts, economic theory
 offers no idea of whether risk
 aversion goes hand in hand with
 patience and other-regarding
 concerns.” (Güth et al. 2008,
 p.261 fn1 omitted)

Behavioral economics is an approach that uses insights from psychology to study some instances of individual behavior that are inconsistent with the predictions and explanations of standard models of the economic agent. The rise of behavioral economics as a prominent approach within the mainstream of economics has been assessed in a number of historical, methodological and philosophical writings.¹ All of these writings focus, by and large, on contributions in behavioral economics that have been published between the end of the 1970s and the first half of the 2000s. This paper explores contributions in behavioral economics that have been published after the middle of the 2000s. The goal is to give a historical and methodological perspective on these contributions so as to characterize a rupture in the evolution of behavioral economics since the middle of the 2000s. Before this date, contributions in behavioral economics consisted, for the most part, in studying behavior *within* either one of three dimensions *separately*, i.e., *either* under risk *or* over time *or* regarding other people. After this date, there has been a growing number of contributions that study behavior *across* these three dimensions, e.g., when risk and time jointly affect individual behavior, or when risk and other people jointly affect individual behavior, and so on.

The historical and methodological perspective taken in this paper focuses on the following normative issue. In classical contributions to behavioral economics, i.e., within the three dimensions separately, standard models were easily used as a normative benchmark to judge observed behaviors as ‘non-rational’ or ‘inconsistent’, notably because they violated some forms of consequentialism. In the new contributions where at least two dimensions affect individual behavior, consequentialist interpretations

¹See, for instance, Sent (2004), Hausman (2011, part III), Angner and Loewenstein (2012), Egidi (2012), Heukelom (2014), Angner (2015), Kao and Velupillai (2015), Grüne-Yanoff (2015), Thaler (2015), Heidl (2016), Geiger (2017).

of standard models loose their normative force so that standard models do not necessarily play the role of normative benchmarks anymore. The following choices have been made to constrain the behavioral economics literature discussed in this paper for space reasons. Firstly, the paper focuses much more on the *qualitative* challenges from behavioral economics than on 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.²

The paper is organized in four section. The first one provides a brief sketch of the classical contributions in behavioral economics prior to the middle of the 2000s, i.e., when the three dimensions were mostly investigated independently of one another. The second one illustrates the contributions in behavioral economics that emerged after the middle of the 2000s where two of the three dimensions are jointly investigated. The theoretical and normative implications of these contributions are then put in historical and methodological perspective in the next two sections. The third section focuses on the normative constraints on decision makers' reasoning across the three dimensions that are imposed by the axioms of separability of the standard models. The fourth section focuses on historically pre-existing arguments against the normative force of the axioms of separability and their implications regarding which of the three dimensions should be taken as the primary one that constrains the other two.

1 Within the three dimensions

This section provides a brief illustration of the type of contributions that are constitutive of the rise of behavioral economics before the middle of the 2000s. The main behavioral regularities (mostly observed in laboratory experiments) studied by behavioral economists are first presented. Then, some

²Game-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).

explanations of the problems they pose for standard models of the economic agent are proposed. Finally, the alternative models proposed in behavioral economics are discussed.

One of the most robust behavioral phenomena that motivated the emergence of behavioral economics 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*. Decision makers face two decision problems. The first one is between a certain consequence (here \$3000) and probable consequences (here \$4000 and \$0), and the observed preference is usually:³

$$\text{The certainty of winning \$3000} \succ 80\% \text{ chance of winning \$4000}$$

The certainty effect is the conjunction of this observation with the one obtained when all the above consequences remain constant but their probabilities are reduced by a common factor (here by four):

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

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*. Decision makers face a first decision problem between consequences occurring now and consequences occurring later, and a second one between the same consequences all delayed later by the same amount of time. Here is an illustration:⁴

$$\begin{aligned} \text{Winning \$5 today} &\succ \text{Winning \$7.50 in a month} \\ \text{Winning \$5 in two weeks} &\prec \text{Winning \$7.50 in a month and two weeks} \end{aligned}$$

Experimental studies of behavior with respect to other people in behavioral economics involve decision problems that have a different empirical structure than the ones in the other two dimensions. The behavioral phenomenon in the dimension of other people with as less strategic interaction as possible is observed in the so-called *dictator game* from Kahneman, Knetsch and Thaler (1986, experiment 2, part 1). Decision makers face only one decision problem, which consists in dividing a sum of money (here \$20) between herself or himself and another anonymous person, and the observed pattern is usually:

³As usual, ‘ \succ ’ means ‘is (strictly) preferred to’. 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).

⁴The illustration comes from the data taken from Faralla et al. (2012, stimuli 107 and 109) because they make the presentation of this effect especially clear. Immediacy and common difference effects have been studied in behavioral economics at least since Thaler (1981), see also Prelec and Loewenstein (1991).

$$\$18 \text{ to self and } \$2 \text{ to other} \prec \$10 \text{ to self and } \$10 \text{ to other}$$

The phenomena illustrated so far are the basic ones in behavioral economics. Conjunctions of these patterns of preferences with further patterns of preferences observed, for instance, when the *signs* of monetary consequences are simply reversed from gains to losses, or when the probabilities or delays are kept constant but the *amount* of money is multiplied by a common factor, create other behavioral phenomena respectively known as *sign effects* and *magnitude effects*. Although it would take too much space to illustrate all these effects, the following point about the ones presented here would apply to sign and magnitude effects just as well: experimental studies in each of the three dimensions implicitly involve the other two. Studies of behavior under risk implicitly assume that the consequences in the decision problems are interpreted as occurring now, i.e., not later, and for oneself, i.e., not for another person. Studies of behavior over time implicitly assume that the consequences in the decision problems are interpreted as being for sure, i.e., not probable, and for oneself. And studies of behavior regarding other people implicitly assume that the consequences in the decision problems are interpreted as being for sure and as occurring now.

These behavioral phenomena violate different types of assumptions of the standard models of the economic agent. For instance, the certainty effect violates *the independence axiom* of expected utility theory, which is necessary for behavior to be represented with standard models of risk preferences. This axiom states that a preference for a lottery (i.e., for a probabilistic distribution of consequences) over another lottery should not change if a common third lottery is added to both of the former lotteries. In the certainty effect, the common third lottery is (implicitly) ‘75% chance of winning nothing’, which, when added to the first pair of lotteries, i.e., ‘the certainty of winning \$3000’ and ‘80% chance of winning \$4000’, gives the second pair of lotteries, i.e., ‘25% chance of winning \$3000’ and ‘20% chance of winning \$4000’, respectively.

The immediacy effect violates the property of *dynamic consistency* that is implied by the axioms of the theoretical framework of exponentially discounted utility, which are necessary for behavior to be represented with standard models of time preferences. Dynamic consistency ensures the consistent evaluation of the same plan (i.e., of the same temporal distribution of consequences) through time, in the sense that its discounted value does not depend on the period from which it is evaluated. In the immediacy effect, it is easy to see that ‘winning \$5 in two weeks’ does not have the same value when it is evaluated now (in which case the agent prefers ‘winning \$7.50 in a month and two weeks’) and when it is evaluated in two weeks, i.e., when ‘two weeks’ becomes ‘today’ (in which case the agent prefers

this plan over ‘winning \$7.50 in a month’).

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. The preference for a fair distribution with another person in the dictator game do not violate an axiom of self-interest because, as Vivian Walsh puts it, “the assumption of self-interest is not presented as a formal axiom [in standard models] – it does not come in, as it were, by the front door” (1996, p.113). Rather, the preference for a fair distribution does contradict a common *interpretation* of standard models as representing self-interested behavior. More precisely, it contradicts the narrowest interpretation of the notion of self-interest, which Amartya Sen calls “self-centered” self-interest, or *self-centeredness* for short, i.e., that “a person’s welfare [or utility] depends only on her own consumption and other features of the richness of her life (without any sympathy or antipathy towards others, and without any procedural concern)” (2002, p.33).

The independence axiom, dynamic consistency and self-centeredness are all normatively justified as necessary components of *rational* behavior 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 verbally 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. 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.

Regardless of the origin of consequentialism in economics, the point is that judging the behavioral phenomena presented at the beginning of this section as non-rational behavior is usually explicitly or

implicitly justified on the ground that these behaviors violate a form of consequentialism. This is most clearly seen in the theoretical work in behavioral economics that tries to account for the behavioral phenomena under risk through prospect theory (Kahneman and Tversky 1979) or cumulative prospect theory (Tversky and Kahneman 1992). Both versions of that theory account for behavioral phenomena such as the certainty effect by using a function that transforms the objective probabilities of a decision problem into the subjective weights that decision makers put on the associated consequences. They also use a specific utility function (the “value function”) that makes risk attitudes dependent on whether consequences are negative or positive in order to account for sign effects. As Peter Wakker (2010, p.162) remarks, this is tantamount to bringing psychologists’ probabilistic sensitivity with economists’ consequentialist sensitivity. This remark has to be understood from a historical perspective: between the 1950s and the 1970s, psychologists working on decision making tended to use only a function on the probability but no function on the consequences and economists working on decision making tended to use only a utility function on the consequences without a function on the probabilities. However, by and large, the dominant position in behavioral economics is that prospect theory only describes behavior but does not allow to judge the described behavior as rational: “traditional EU [expected utility] is, in my opinion, the hallmark of rationality, any deviation from final wealth [i.e., consequentialism] 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).

Models of quasi-hyperbolic discounting (esp. Laibson 1994; 1997) – inspired by psychiatrist George Ainslie’s (1975; 1992) theory of hyperbolic discounting – play a role for time preferences in behavioral economics comparable to Kahneman and Tversky’s prospect theory for risk preferences. Both quasi-hyperbolic and hyperbolic models of discounting use a function to discount time in a way that allows to represent intertemporal preference reversals, i.e., violations of dynamic consistency. These models are formalized with the use of game theory and interpreted as different (multiple) selves of a given individual evaluating the same plan at different points in time. In both cases, the discount rate generated by the model is not constant, but depends on the period at which a plan is evaluated. Consequences that are far from occurring now are much more discounted than consequences that are closed to occurring now (this tendency is smooth in hyperbolic discounting models while it is binary in quasi-hyperbolic ones). As with expected utility theory (though somewhat less clearly), there is a general tendency shared by both standard and behavioral economists to use exponential discounting as a normative benchmark to judge violations of dynamic consistency, i.e., behavior generated by (quasi-)

hyperbolic discounting, as non-rational (see, e.g., Gollier 2001; Loewenstein et al. 2015, pp.61-65).

By contrast with behavioral models of risk and time preferences, there is not one but several – at least five – theoretical alternatives for social preferences in behavioral economics (Rabin 1993; Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Andreoni and Miller 2002; Charness and Rabin 2002). All these different ways to introduce non-self-centered motives in utility functions can be interpreted 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 person gets more or less than the decision maker (Fehr and Schmidt 1999), (2) on whether the other persons get on average more or less than the decision maker (Bolton and Ockenfels 2000), (3) on the other person(s) who get(s) the least or on (4) how much the decision maker *and* the other person (as a group) get in total (both Andreoni and Miller 2002 and Charness and Rabin 2002 propose characterizations (2) and (3)). The decision maker’s utility can also depend on (5) the other person’s (a) own motives and (b) beliefs about the decision maker’s motives (Rabin 1993). Despite these differences, all these models represent a decision maker who is not *only* maximizing her or his own welfare, though they all represent a decision maker who is nevertheless *also* maximizing his or her own welfare. This explains why these modeling of violations of self-centeredness are always accompanied with statements about the rationality of the underlying behavior. Hence there is a sharp contrast with the other two dimensions, i.e., with violations of the independence axiom or dynamic consistency.

2 Behavioral phenomena across the three dimensions

Studies of behavior at the intersection of at least two dimensions do not references each other very much. The goal of this section is to provide a general picture of these new behavioral phenomena that allows to appreciate the depth of this new trend in behavioral economics that characterizes its development after the middle of the 2000s. For historical reasons pertaining to the empirical structure of the classical challenges discussed in the previous section, it is possible to flesh out one feature of the ‘logic’ of interactions across dimensions, namely that there are *directions of introduction of one dimension in another one*.

For instance, Manuel Baucells and Franz Heukamp (2010, table 1) study the interaction between risk and time by *introducing time in the dimension of risk*. That is, they first replicate a classical certainty effect with however the dimension of time made explicit, i.e., they add ‘now’ to all the

consequences:

$$\begin{aligned} \text{The certainty of winning €9 now} &> 80\% \text{ chance of winning €12 now} \\ 10\% \text{ chance of winning €9 now} &< 8\% \text{ chance of winning €12 now} \end{aligned}$$

Then, they use the same decision problems but delay all the consequences by the same amount of time (three months) and observe the following pattern of preference:

$$\begin{aligned} \text{The certainty of winning €9 in 3 months} &< 80\% \text{ chance of winning €12 in 3 months} \\ 10\% \text{ chance of winning €9 in 3 months} &< 8\% \text{ chance of winning €12 in 3 months} \end{aligned}$$

In other words, they observe that introducing time in the dimension of risk in this way can restore consistency in risk preferences so that the behavior can be represented in standard models. One of the main inspirations of Baucells and Heukamp's experiment was a series of experimental results obtained fifteen years earlier by psychologists Gideon Keren and Peter Roelofsma (1995). The latter paper is indeed one of the main trigger (with ten years of delay) of the mid-2000s emergence of interests in the interactions between time and risk preferences in behavioral economics. Keren and Roelofsma (1995, exp.1) also *introduced risk in the dimension of time*. That is, they first replicate a classical immediacy effect with however the dimension of risk made explicit, i.e., they add that the consequences occur with certainty:

$$\begin{aligned} \text{The certainty of winning €100 now} &> \text{The certainty of winning €110 in 4 weeks} \\ \text{The certainty of winning €100 in 26 weeks} &< \text{The certainty of winning €110 in 30 weeks} \end{aligned}$$

Then, they use the same decision problems but add an equal risk (50% chance) to all the consequences and observe the following pattern of preference:

$$\begin{aligned} 50\% \text{ chance of winning €100 now} &< 50\% \text{ chance of winning €110 in 4 weeks} \\ 50\% \text{ chance of winning €100 in 26 weeks} &< 50\% \text{ chance of winning €110 in 30 weeks} \end{aligned}$$

In other words, they observe that introducing risk in the dimension of time in this way can restore consistency in time preferences so that the behavior can be represented in standard models. Hence the interactions between risk and time preferences can go both ways: time preferences can restore consistency in risk preferences (Baucells and Heukamp 2010; Keren and Roelofsma 1995, exp.2), and risk preferences can restore consistency in time preferences (Keren and Roelofsma 1995, exp.1). These two sets of phenomena are often discussed altogether in the post mid-2000s literature on interactions across dimensions, i.e., apart from other interactions involving the dimension of other people.

An instance of the *introduction of risk in the dimension of other people* is the dictator games of Michal Krawczyk and Fabrice Le Lec (2010). They construct a “probabilistic dictator game” (p.500) where a decision maker has to allocate 10 tokens between herself or himself and another person in different conditions. In a deterministic condition mimicking standard dictator game under certainty, one token is worth 10% of the consequences of a predetermined allocation, e.g., if the predetermined allocation is €20 for the decision maker and €20 for the other, then keeping the 10 tokens implies €20 for the decision maker (100% of his consequences) and nothing for the other (0% of his consequences), while giving 5 tokens to the others implies €10 for the decision maker (50% of his consequences) and €10 for the others (50% of his consequences). In another condition with risk, one token is worth 10% *chance* of winning the consequences of a predetermined allocation, e.g., (with the same €20/€20 allocation) keeping 10 tokens implies the certainty (100% chance) of winning €20 for the decision maker and the impossibility (0% chance) of winning €20 for the other, while giving 5 tokens to the other implies a lottery with 50% chance of winning €20 and 50% chance of winning nothing for the decision maker, and the same lottery for the other. The decision makers who gave at least one token in the first case (they gave 2.43 on average) gave significantly less token in the second case (they gave 1.98 on average). With other results from other conditions, the observed pattern of preference is that introducing risk in this inter-personal decision problem in this way tends to make decision maker behavior self-centereded. That is, the introduction of risk makes behavior closer to the narrowest interpretation of self-interest with respect to standard models.

An instance of the *introduction of time in the dimension of other people* is in the dictator games of Jaromir Kovarik (2009). He constructs dictator games where the decision maker has to allocate €6 between himself and the other person in different conditions. In an immediate condition mimicking the standard dictator, the decision maker and the other both receive their money at the end of the experiment. The other conditions vary the delay with which the money is received, from two days to twenty two days. The observed pattern of preference has some similarity with the previous one: introducing time in this inter-personal decision problem in this way tends to make decision maker behavior self-centered, with virtually everybody being strictly self-centered after fourteen days of delay (i.e., all give €0 between fourteen twenty two days of delay). Hence introducing risk and time in the dimension of other people *can* push decision makers to abide by the narrow interpretation of self-interest (self-centeredness) that was initially violated in the dictator game.

We can finally turn to how other people can be introduced in the dimensions of risk and time.

An instance of the *introduction of other people in the dimension of risk* is in the experiments of Gary Bolton and Axel Ockenfels (2010), with decision problems and patterns of preference such as the following:

The certainty of winning €9 for me	>	50% chance of winning €16 for me <i>and</i> 50% chance of winning nothing for me
The certainty of winning €9 for me and €16 for the other	<	50% chance of winning €16 for me and €16 for the other <i>and</i> 50% chance of winning nothing for me and nothing for the other

Hence, decision makers exhibit a risk preference for oneself (i.e., risk aversion) in the first decision problem that reverses itself (for risk seeking) when the other person is introduced in the second decision problem. Notice that, by contrast with the previous examples where consistency with respect to standard models was restored by the introduction of another dimension, here the introduction of another dimension creates a new inconsistencies with respect to standard models. Indeed, consequentialism is clearly violated here as the consequences for oneself are invariant across the two decision problems. The second pair just adds inequality in the sure object of choice and equality in the risky one. Thus it seems that a preference for equality (or against inequality) interacts with risk preferences, creating inconsistency in the latter (at least with respect to standard models).

To my knowledge, *there is no experiments introducing other people in the dimension of time*, at least in the same fashion as the experiments discussed so far in this subsection, i.e., where the introduction of other people in an intertemporal choice would create a conflict that would have an effect on time preferences⁵. This is surprising for two reasons. First, such decision problems are not hard to imagine in an experimentally implementable way. For instance, one instance drawing on the previous choice experiment would look like this:

€9 for me now	vs.	€16 for me in a week
€9 for me now and €16 for the other now	vs.	€16 me in a week and €16 for the other in a week

Second, historically, since at least the first quarter of the twentieth century, there has been a substantial literature on the relation between discounting and altruism in macroeconomics, related to the

⁵This passage was written in 2016. Since then the literature on interactions across dimensions has continued to grow exponentially and such examples exist. The next version of this paper will very soon integrate these new contributions.

problem of intergenerational altruism. This literature has of course grown more rapidly in the last couple of decades because of the issues involved in climate change. In this movement, some macroeconomists interested ‘social discounting’ have started to look at the behavioral economics (see the references in Peroco and Nijkamp 2009). An important problem related to individual time preferences in social discounting is illustrated by Sen’s (1961, pp.487-9) “*Isolation Paradox*”, which could also be of interest for experiments in behavioral economists. The original thought experiment runs roughly as follows. A decision maker who will be dead in twenty years prefers one unit of consumption now to three units of consumption in twenty years, even though he or she cares about the future generations, i.e., three for the (later) other is not enough for her or him to sacrifice one for himself or herself. Another individual confronted with the same problem comes to the decision maker and says: ‘if you save your three units, then I will do the same, if you consume your unit now, then I will also do the same’. The decision maker changes his or her preferences, and Sen argues that this reversal is perfectly rational, even without deep moral considerations, i.e., only on quantitative grounds, sacrificing one unit for oneself for six units for the others is enough. This would easily translate in the following decision problems with the following pattern of preferences:

$$\begin{aligned} \text{€10 for me now} & \succ \text{€30 for a group of people who need money in a week} \\ \text{€10 for me now and €10 for another person now} & \prec \text{€60 for a group of people who need money in a week} \end{aligned}$$

Summing up, interactions across dimensions can (with respect to the standard models) either restore consistency in risk, time or social preferences, or produce further inconsistencies. Hence the patterns of preferences presented in this section are, taken altogether, in contradictions with *both* the standard model *and the behavioral alternatives* presented in the previous section.

3 Crosscutting separability

In standard models, the implicit connections between the three dimensions are due to mathematical results and theoretical contributions of John Harsanyi (1955), William Gorman (1968) and Gérard Debreu (1960). We shall explain these connections by following one of the rare discussion of the implicit connections between the standard models of the three dimensions taken altogether: John Broome’s (1991) book-length investigation of the theoretical and normative interpretations of these contributions. Broome’s account focuses on the role of *separability* in the construction of functional representation ($U(\cdot)$) of preferences (\succsim). We shall first illustrate that the standard models studied

above share a common requirement of separability, namely *additive separability*, before we illustrate Broome’s account of the interactions across dimensions through what he calls *crosscutting separability*.

In the dimension of risk, preferences are separable between mutually-exclusive states, e.g., in a coin flip lottery (X), there are only two mutually exclusive states, namely of head *or* tail coming up (1 *or* 2, respectively), the consequences in each being therefore mutually exclusive as well (you get x_1 *or* x_2). Additive separability requires, not only that each of these consequences are evaluated separately, but that they can be added, i.e., $U(x_1, x_2) = u_1(x_1) + u_2(x_2)$. The independence axiom in expected utility theory is also separability condition that requires the utilities of consequences to be weighted linearly by their respective probabilities, i.e., $U(X) = .5u_1(x_1) + .5u_2(x_2)$. Furthermore, it is common to assume a unique utility function for the consequences, i.e., $U(X) = .5u(x_1) + .5u(x_2)$. In the dimension of time, preferences are separable between mutually exclusive periods of a plan (X), e.g., now *or* later (1 *or* 2, respectively), with mutually exclusive consequences x_1 *or* x_2 . What is sometimes called ‘time separability’ implies additive separability, i.e., $U(X) = u_1(x_1) + u_2(x_2)$. Time separability is part of the axioms necessary for exponential discounting, which together imply the weighting of the utilities of the consequences to be an exponential function of time, i.e., $U(X) = u_1(x_2) + \frac{u_2(x_2)}{1+\rho}$, with ρ a constant discount rate. Again, it is common to assume a unique utility function for the consequences, i.e., $U(X) = u(x_2) + \frac{u(x_2)}{1+\rho}$. In the dimension of other people, preferences are separable between the mutually the exclusive people that are the recipients of an allocation (X), e.g., me and you (1 and 2, respectively) receives consequences x_1 and x_2 that are mutually exclusive so that $X = x_1 + x_2$. In modeling whoever evaluates an allocation (e.g., me, you, or a social observer), additive separability between people is often implied by axioms on preferences or assumed directly, i.e., $U(X) = u_1(x_1) + u_2(x_2)$, together with various and usually controversial assumptions or axioms about how to weight and/or compare the utilities of people (see Sen 1977a; or Harsanyi 1988 on these issues).

Crosscutting separability concerns preferences over consequences that are distributed across dimensions. To illustrate what crosscutting separability means and requires, as well as the analytical connection it gives rise to, we shall use Broome’s type of example that are, as he puts it, “conveniently symmetrical” (1991, p.62) on two dimensions only. To do so, consider the following ‘temporal allocation’ or ‘social plan’:

$$\begin{array}{ccc} \text{€10 now and €10 soon for me, and} & & \text{€5 now and €20 soon for me, and} \\ \text{€10 now and €10 soon for you} & \text{vs.} & \text{€20 now and €5 soon for you} \end{array}$$

This decision problem can be represented as in Table 1, where the subscripts separated by a comma

		Time			
		Now	Soon	Now	Soon
People	Me	€10 ($a_{1,1}$)	€10 ($a_{2,1}$)	€5 ($b_{1,1}$)	€20 ($b_{2,1}$)
	You	€10 ($a_{1,2}$)	€10 ($a_{2,2}$)	€20 ($b_{1,2}$)	€5 ($b_{2,2}$)
		A		B	

Table 1: A conveniently symmetrical example for crosscutting separability

indicate the coordinates of what Broome calls the “locations”. The first one is the time coordinate, i.e., location 1 is now and location 2 is soon, the second one is the people coordinate, i.e., location 1 is me and location 2 is you.

Crosscutting separability implies that you can evaluate the row and the column of this table separately. Because crosscutting separability implies additive separability, there is at least four utility functions evaluating the consequences here, i.e., $u_{1,1}(\cdot)$, $u_{2,1}(\cdot)$, $u_{1,2}(\cdot)$, $u_{2,2}(\cdot)$. Connections across dimensions occur because one evaluation on one dimension will depend on my preferences regarding the other dimension. This is the case because each consequence cuts across *two* dimensions (i.e., is localized in one location of each dimension). Hence, each consequence is evaluated by *one* utility function, which captures both time and social preferences, which are thus *identical*. Separability conditions are often considered to be technical assumptions without empirical or normative contents (see, e.g., Gilboa 2009, p.52). But somewhat paradoxically, experimental results from behavioral economics are often taken to demonstrate empirical violations of such separability conditions. Most of the empirical results presented so far illustrate such empirical violations. The following illustrates how separability imposes normative constraints on the reasoning that decision makers ought to perform to be considered as rational reasoners and how that form of rationality applied across more than one dimensions implies that risk, time and social preferences should be identical. For ease of presentation, that reasoning is presented at the first person.

Evaluating time locations (i.e., the columns) separately means that I can subevaluate what happens now separately from what happens soon. When I evaluate what happens now, I see that if I choose A , you and me get the same thing (€10), and if I choose B , we do not get the same thing (I get €5 and you get €20). On this *time* location, my choice therefore depends on (at least one aspect of) my *social* preferences: what are my attitudes towards an *equal* treatment in which both of us receive the same thing, and towards an *unequal* treatment in which there is €20 for you and €5 for me? If I like equality, I will tend to choose A , if I don’t mind about inequality (and/or if I genuinely care about

other people at the expense of my self-centered self-interest) I will tend to choose *B*. When I evaluate what happens soon, I see that if I choose *A*, you and me both get the same thing (€10), and if I choose *B* we do not get the same thing (I get €20 and you get €5). Again, on this time location my choice depends again on my social preferences: if I like equality I still tend to choose *A* and if I do not mind about inequality (and/or if I genuinely care only about me), I will still tend to choose *B*. Therefore, evaluated separately on the dimension of time, the result is the same whatever the time locations: if I like equality, I choose *A*, if I don't mind about inequality, I choose *B*. Thus, the result on the dimension of time depends on (at least one aspect of) my social preferences.

Evaluating people locations separately means that I can evaluate what happens for me separately from what happens for you. When I evaluate what happens for me, I see that if I choose *A* I get €10 now and €10 soon, and if I choose *B*, I get €5 now and €20 soon. On this *personal* location, my choice therefore depends on (at least one aspect of) my *time* preferences *which does not correspond directly to discounting but to the elasticity of intertemporal substitution*: what are my attitudes towards an equally distributed flow of money between two points in time separated by a week, and towards an unequally distributed flow of a slightly bigger amount of money between the same points in time? If I like steady flows of money (where I get the same amount periodically), I will tend to choose *A*, but if I like unsteady flows of money (where I get more at some points though less at others), I will tend to choose *B*. When I evaluate what happens for you, I see that if I choose *A*, you get €10 now and €10 soon, and if I choose *B*, you get €20 now and €5 soon. Again my choice on this location depends on (at least one aspect of) my time preferences: if I like equally distributed flow of money I will tend to choose *A*, but if I like unequally distributed flows of money I will tend to choose *B*. Therefore, evaluated separately on the dimension of people, the result is the same whatever the people locations: if I like equally distributed flow of money, I will tend to choose *A*, but if I like unequally distributed flows of money I will tend to choose *B*. Thus, the result on the dimension of people depends on (at least one aspect of) my time preferences. Notice that high discounting leads to a preference for plans with distribution of consequences skewed towards the present. By contrast, high elasticity of intertemporal substitution leads to a preference for plans with unequal distribution of consequences, not necessarily skewed, but usually assumed to be so towards the future because of positive interest (or growth) rates. Hence, from a revealed preference perspective, violations of dynamic consistency *can* come from a change in the former *and/or* the latter. Probably because the latter is meaningful in economics only with respect to further economic variables that are absent from experimental setups, measures

derived from intertemporal experiments are interpreted as being about ‘*pure*’ time preferences, i.e., discounting. Nevertheless, both parameters can non-controversially said to constitute decision makers’ attitudes over time. As Frederick, Loewenstein and O’Donoghue (2002, p.359) put it:

“diminishing marginal utility (that the instantaneous utility function [...] is concave) and positive time preference (that the discount rate ρ is positive) [...] create opposing forces in intertemporal choice: diminishing marginal utility motivates a person to spread consumption over time, while positive time preference motivates a person to concentrate consumption in the present.”

When we bring the evaluation on the dimension of time and the evaluation on the dimension of people together, we have a very simple and elegant conclusion. On the one hand, if I choose *A*, it reveals a preference for equality in the distribution of money across people, and a preference for equality in the distribution of money across time. On the other hand, if I choose *B*, it reveals a preference for inequality in the distribution of money across people, and a preference for inequality in the distribution of money across time. Hence disliking inequality in the distribution of money over time *and* across people “must be linked”: “[*t*]/*hey* are really just two ways of describing the same pattern of preferences” (Broome 1991, pp. 63-64, my emphasis). Adding one dimension to the previous reasoning generalizes it, the point being that, in standard models, *one* consequence that cuts across *two or three dimensions* is evaluated by *one* single utility function that identifies aspects of risk, time and social preferences with one another. The plausibility of the argument depends on crosscutting separability and “[i]f separability fails in one or the other dimension, the conclusion will fail too” (ibid). Of course, violations of the independence axiom, dynamic consistency and self-centeredness imply that separability tends to fail empirically in each dimension. Broome is aware of this but his perspective is normative – and related to discussions of decisions about future unknown social positions under a ‘veil of ignorance’ where decision makers’ risk and inequality aversion *ought* to be identical. Roughly, he argues from a version of consequentialism that violations of separability in the dimension of risk and other people are irrational; while those in the dimension of time are rationally justifiable because separability in that dimension carries a metaphysical implication of multiple-selves which he finds unappealing.

Some empirical implications of the theoretical issues discussed by Broome have recently resurfaced in the normative economics of climate change, especially in a paper by Giles Atkinson, Simon Dietz, Jennifer Helgeson, Cameron Hepburn and Hakon Soelen (2009). One such implication is that standard social welfare functionals can be calibrated with data about either one of these three dimensions (i.e., risk aversion, elasticity of intertemporal substitution and inequality aversion), notably to calculate

the social *discount rate* through which policies are evaluated. But in practice these are not identical because decision makers usually violate separability conditions. Hence calibrations with either one of the three types of data imply different policy recommendations from calibrations with either one of the two others. Thus, “[a]n important theoretical lacuna still exists because no model to date enables all three concepts to be disentangled simultaneously” (p.3). Furthermore, they argue against normative defense à la Broome for the identification of even two of the three aspects. Such a defense, on their account, presupposes conditions for reasoning that are unrealizable in practice (i.e., veil of ignorance-like). They further argue that normative justifications for the articulation (or lack thereof) among attitudes in the three dimensions need to be debatable in a democratic decision process, hence possibly going against consequentialism. Accordingly, they conduct a large scale internet survey to elicit these three attitudes of people regarding climate change issues, and found large differences and furthermore weak correlations among the three, contrary to “the standard welfare model [which] implicitly assumes a perfect correlation” (p.4). One implication is that an extra value judgment is needed in applied work, namely, from which of the three dimensions should we take the data to calibrate social welfare functionals? In other words, which of the three dimensions should constrain the other two?

Finally, the paper by Güth, Vittoria Levati and Matteo Ploner (2008) quoted in the epigraph to this paper tackles such correlations across the three dimensions using lab experiments. Decision makers are asked the minimum amount of money they would accept to sell each one of these objects of choice (i.e., their willingness to accept, WTA). Here are four examples of the sixteen objects of choices, where the lottery gives 50% chance of winning €16 and 50% chance of winning €38, ‘later’ is in three months, and ‘you’ is the decision maker (ibid, appendix, emphasizes and bolded terms are theirs):

You get €27 for sure *now*, and **the other** gets the lottery *now*.
You get €27 for sure *later*, and **the other** gets the lottery *later*.
You get €27 for sure *now*, and **the other** gets the lottery *later*.
You get €27 for sure *later*, and **the other** gets the lottery *now*.

The main tendencies are that decision makers care about the other’s consequences when their own consequences are immediate and sure, that their discount rates reveal much more impatience for their own consequences than for the other’s, and that risk aversion usually goes with more impatience and risk seeking with more patience.

Notice that empirical studies traditionally investigate correlations between, on the one hand, either risk *or* time *or* social preferences, and, on the other hand, something ‘external’ to these preferences,

e.g., some measures of economic success (as in Tanaka et al. 2010) or some measures of stability over the years in a given data panel of one type of preferences (as in Chuang and Schechtner 2015). By contrast, both Atkinson et al. (2009) and Güth et al. (2008) study ‘internal’ correlations among the three types of preferences. This approach is less traditional though not new *per se* (see the references cited by Jamison et al. 2012, p.14), but it can be argued that it is a growing trend motivated by behavioral economics (see esp. Dean and Ortoleva 2015). Two remarks about this trend are worth making, as they respectively introduce the objects of the next two subsections. The first one is that this trend is globally silent on the normative dimension (i.e., à la Broome) or implications (i.e., à la Atkinson et al.) of their empirical results. The second one is that, though helpful to understand some underlying issues of the empirical results presented in the previous section, quantitative correlations among dimensions are different from the qualitative issue of how new inconsistencies arise or disappear with the introduction of new dimensions into a given one. Notably, quantitative correlations do not directly address how to get a parsimonious theoretical account of the patterns of preferences displayed in the two previous sections taken altogether. Hence it could be argued that without such a qualitative work available, it is impossible to even represent the quantitative variations in risk time and/or social preferences underlying the choices of *one* decision maker abiding by the regularities of these two pictures, let alone investigate the correlations among them.

4 Arguments for the primacy of one dimension over the others

The goal of this section is to provide further illustrations of the problematic implications of separability across dimensions that were historically posed before the middle of the 2000s. Three illustrations have been selected because they deliver some insights about the normative issue of which of the three type of preferences should constrain the other two. The first two examples provide arguments for a primacy of time preferences and the last one provide arguments for a primacy of risk preferences.

The first illustration is about the so-called issue of ‘the timing of uncertainty resolution’. The classical paper stating the issue is David Kreps and Evan Porteus’ (1978), who use the following intuitive example (p.185). Consider the coin flip: if heads comes up you win €5 now and €10 later, if tail comes up you win €5 now and nothing later. Because there is a common consequence for now (€5), expected utility implies that the decision makers should be *indifferent* between flipping the coin now – so that the uncertainty about whether you get €10 or nothing later resolves now – or flipping it later – i.e., you get your €5 now anyway but the uncertainty about whether you get €10 *or* nothing later

resolves later. Kreps and Porteus' argument is twofold. On the one hand, they argue that there are good reasons for not being indifferent, notably a preference for early resolution can be justified from the ease it creates for budgeting now one's future expenditures. On the other hand, they propose an axiomatic framework from which models that analytically distinguish risk and time attitudes can be constructed. While such models have had their success in macroeconomics (see Gollier 2001, chap.20), they have not been discussed much in behavioral economics (though see Coble and Lusk 2010 and the reference therein for contributions in experimental economics).

The second illustration consists in pointing some conflicts between consequentialism and dynamic consistency. Originally, the first instances of these challenges have been raised in normative economics (e.g., Diamond 1967; or more recently, Giraud and Renouard 2011; Bovens 2015; and more generally, see Mongin and d'Aspremont 1998). The structure of the argument has been applied to decision theory and Hammond's version of consequentialism in a classical paper by Mark Machina (1989), who use the following intuitive example (pp.1643-4). Mom has an indivisible treat, say one candy, she can give to *either* one of her children, Abigail *or* Benjamin. Mom is indifferent between (A) 'Abigail getting the candy for sure' and (B) 'Benjamin getting the candy for sure', and strongly prefers either one of these consequences to the one whereby (C) 'None of them gets the candy for sure'. Mom however strictly prefers the probabilistic consequences – she can construct by flipping a coin – of ($D = (A, .5; B, .5)$) '50% chance that Abigail gets the candy and 50% chance that Benjamin gets the candy' to either one of the sure consequences. Formally, Mom's preferences are $D \succ A \sim B \succ C$. So Mom flips a coin, and it turns out that Abigail wins the candy. But Benjamin steps in and says: 'Mom, you told us earlier that you preferred 'to flip a coin' over 'Abigail getting the candy for sure' ($D \succ A$), *now* Abigail is getting the candy for sure so please respect your preferences and flip a coin'. In other words, Benjamin argues from (Hammond's) consequentialism that Mom should respect the independence axiom of expected utility by sticking to the part of her initial preferences that does not violate the requirement of linearity in the weighting of probabilities (i.e., preferring the gamble over *one* sure consequence, $D \succ A$). Mom answers "You had your chance!" and gives the candy to Abigail. In other words, Mom argues that she is happy to violate consequentialism because she wants to be dynamically consistent with the part of her initial preferences that violates expected utility (i.e., preferring the gamble over the indifference between *both* sure consequences, $(A, .5; B, .5) = D \succ A \sim B$).

Machina (1989) argues at length that (Hammond's) consequentialism is a dynamic version of the very separability requirement that decision makers usually violate in the first place. Hence, if they

have normative justifications for their violations, which is clearly the case for Mom here, another type of dynamic consistency than the one implied by Hammond’s consequentialism is needed for these decision makers. In Mom’s example, the reason justifying her choice is a preference for fairness, indeed following the original literature in normative economics where these counter-examples to the normative implications of expected utility originally emerged. But Machina’s point is more general. Roughly put, it is that ‘history matters’ in the following sense. Mom’s preferences for ‘giving the candy to Abigail *if* the uncertainty of 50% chance that Benjamin Benjamin gets the candy *has been borne* (but not realized)’ is already her preferences *ex ante*, i.e., before she knows the outcome of the coin flip. Hence by saying ‘no’ to Benjamin *ex post* she is being dynamically consistent with her *ex ante* preferences.

Reading Machina’s paper through the lenses of interactions across the three dimensions, which is *not* the theme of his paper, one realizes how much interactions across the three dimensions are *sources of reasons* – or “*generators of reasons*” (Broome 1991, p.23, my emphasis) – in economics. For economic agents faced with decision conflicts in one dimension, e.g., risk in Mom’s case, another dimension is used to bring the justification that resolves the conflict, e.g., other people in Mom’s case. For economists intending to formally represent such resolution in one dimension, contributions from other economists working in subfields that focus on other dimensions become relevant. In Machina’s case, it is not only the contributions from normative economics in the dimension of other people, but also from intertemporal consumption, as he makes an analogy with the non-separable time preferences used there to ground his arguments (1989, pp.1644-1645). Interactions across the three dimensions are also sources of reasons in the sense that they make intuitive decision theoretic arguments *appealing* in economics. Take for instance the well-known Dutch book (under uncertainty) or, more generally (under certainty) money pump arguments that are used to establish the *economic* rationality of not violating the independence axiom or the axiom of transitivity. The appeal of these arguments relies on a person who has rather *anti-social preferences* toward the decision maker, as the former tries to put the latter to ruin *over time* (by repeated decisions).

Finally, notice that the structure of the interactions across dimensions in both Kreps and Porteus’s and Machina’s papers is not the same as the one in the empirical results presented in the previous sections. Furthermore, both papers have been relatively disconnected from behavioral economics before the middle of the 2000s. However, the theoretical outcomes of the empirical results presented in the previous section have made such a connection. This is so for Kreps and Porteus because Thomas Epper and Helga Fehr-Duda’s (2015) theoretical unification of interactions across risk and

time includes the former’s phenomenon as one to be unified among the other ones (even if the other ones are observed through laboratory experiments and the former is an intuitive example). And this is so for Machina because both Fudenberg and Levine’s (2012a) and Stefan Trautmann and Wakker’s (2010) theoretical contributions motivated by the results from the previous section (respectively about the interactions across risk and other people, and across other people and time) make non-trivial uses of Machina’s reasoning. In a nutshell, Fudenberg and Levine show that all the models of non-self-centered motives discussed in the first section cannot be extended under expected utility theory without violating a preference for *ex ante* fairness. That partly explains why none of these models can account for interactions across the dimensions of risk and other people presented in the previous section. Trautmann and Wakker make roughly the same point and emphasize how the phenomenon is better understood by connecting it with violations of dynamic consistency in the dimension of time. Finally, Andreoni et al. (2016) proposes a set of experiments to test whether decision makers tend to be consequentialist or dynamically consistent in social allocation problems under risk. The answer is that it depends, among other factors, on whether they make their decisions *ex ante* (in which case they tend to be dynamically consistent) or *ex post* (in which case they tend to be consequentialist).⁶

The last illustration comes from the work of psychologists Keren and Roelofsma (1995) and has been formalized in economics by Yoram Halevy (2008). Keren and Roelofsma argue that the future is *necessarily* risky for various reasons but at least because of *the omnipresent possibility of sudden death*. On the other hand, risk can also be in the present which is thus *not necessarily* certain. Furthermore, the inherent riskiness of consequences in the future stemming from the possibility of sudden death can only disappear through the immediate experience of those consequences. Halevy accounts for Keren and Roelofsma’s (1995) results (presented in the second section of this paper) about the introduction of time in the dimension of risk and *vice-versa*. Halevy does not just weaken some axioms of expected utility theory or underlying exponential discounting to be compatible with the psychologists’ results. His modeling strategy sought to be compatible with their explanations from the omnipresent possibility of sudden death. To do so, Halevy model the future “as a random process that has a positive probability of stopping at any given period” and “may be interpreted as the hazard of mortality” (2008, p.1145). This captures the implicit risk inherent in the future, which is absent

⁶Rotemberg (2014, sect.6) provides a recent review of the post mid-2000s’ interactions across the dimension of risk and social preferences. Trautmann and Vieider (2012) provide a very wide survey of the different ways by which other people can be introduced in the dimension of risk in behavioral economics and social psychology. It should also be noted that Fudenberg and Levine’s (2012a) contribution has motivated further empirical (see esp. Brock et al. 2013) and theoretical (see esp. Saito 2013) contributions on the interactions between risk and social preferences.

in the present (i.e., when this probability equals one) so that the non-linear weighting function that transforms the probability gives disproportionate weights to the present, inducing a preference for the present and dynamic inconsistency. The model nonetheless uses an exponential discount function, hence not the whole of time discounting stems from risk preferences. Thus, Halevy’s model can account for the facts that quasi-hyperbolic models were designed to account for as well as the facts that are in contradiction with these models, i.e., Keren and Roelofsma’s disappearance of the immediacy effect when the present is uncertain.⁷

Conclusion

As a conclusion, here are three remarks about the comparison between the classical behavioral economics contributions within the three dimensions and the recent ones across the three dimensions. Firstly, neither standard nor behavioral models can account for the set of behavioral phenomena that occur both within and across dimensions. Hence trying to account for both is a potentially fruitful area of theoretical research for those who seek to reconcile standard and behavioral economics. Secondly, it is more problematic to use standard models as normative benchmarks to judge behavior as rational or non-rational when there are interactions between at least two dimensions than when the dimensions are investigated independently. This provides an opportunity to discuss the relevance of other forms of rationality than the consequentialist one that is traditionally used in economics. Thirdly, it can be argued that all significant decisions in the real world involves the three dimensions of rationality, not just two. Hence investigating jointly the interactions of the three dimensions altogether can be a fruitful area of empirical research for those who are interested in ecological validity.

References

- 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.
- Andreoni, J., Aydin, D., Barton, B., Bernheim, D. B. and Naecker, J.: 2016, When Fair Isn’t Fair:

⁷Saito (2011) provides a non-trivial technical correction to Halévy (2008).

- Sophisticated Time Inconsistency in Social Preferences, *Working Paper* .
URL: http://web.stanford.edu/daydin/DAYdin_Fair.pdf
- 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.
- 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.
- 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.
- Baucells, M. and Heukamp, F. H.: 2010, Common Ratio Using Delay, *Theory and Decision* **68**(1–2), 149–158.
- Baucells, M. and Heukamp, F. H.: 2012, Probability and Time Trade-off, *Management Science* **58**(4), 831–842.
- 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.
- Bovens, L.: 2015, Evaluating Risky Prospects: the Distribution View, *Analysis* **75**(2), 243–253.
- Brock, M. J., Lange, A. and Ozbay, E. Y.: 2013, Dictating the Risk: Experimental Evidence on Giving in Risky Environments, *The American Economic Review* **103**(1), 415–437.
- Broome, J.: 1991, *Weighing Goods: Equality, Uncertainty and Time*, Basil Blackwell, Oxford.

- Camerer, C. F.: 2003, *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, Princeton.
- Charness, G. and Rabin, M.: 2002, Understanding Social Preferences with Simple Tests, *The Quarterly Journal of Economics* **117**(3), 817–869.
- Chuang, Y. and Schechter, L.: 2015, Stability of Experimental and Survey Measures of Risk, Time, and Social Preferences: A Review and Some New Results, *Journal of Development Economics* **117**(November), 151–170.
- Coble, K. H. and Lusk, J. L.: 2010, At the Nexus of Risk and Time Preferences: An Experimental Investigation, *Journal of Risk and Uncertainty* **41**(1), 67–79.
- Dean, M. and Ortoleva, P.: 2015, Is it All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena, *Working Paper* .
URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2643355
- Debreu, G.: 1960, Topological Methods in Cardinal Utility Theory, in K. J. Arrow, S. Karlin and P. Suppes (eds), *Mathematical Methods in the Social Sciences*, Stanford University Press, Stanford, pp. 16–26.
- Diamond, P. A.: 1967, Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility, *The Journal of Political Economy* **75**(5), 765–766.
- 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.
- 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
- 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-Duda, H. and Epper, T.: 2012, Probability and Risk: Foundations and Economic Implications of Probability- Dependent Risk Preferences, *Annual Review of Economics* **4**(1), 567–593.

- Fehr, E. and Schmidt, K. M.: 1999, A Theory of Fairness, Competition, and Cooperation, *The Quarterly Journal of Economics* **114**(3), 817–868.
- 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. and Levine, D. K.: 2012, Fairness, Risk Preferences and Independence: Impossibility Theorems, *Journal of Economic Behavior & Organization* **81**(2), 606–612.
- 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.
- Geiger, N.: 2017, The rise of behavioral economics: A quantitative assessment, *Social Science History* **41**(3), 555–583.
- Giraud, G. and Renouard, C.: 2011, Is the Veil of Ignorance Transparent?, *Oeconomia* **1**(2), 239–258.
- 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.
- 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.
- 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.: 2011, *Preference, Value, Choice, and Welfare*, Cambridge University Press, New York.
- Heidl, S.: 2016, *Philosophical Problems of Behavioural Economics*, Routledge, London and New York.
- Heukelom, F.: 2014, *Behavioral Economics: A history*, Oxford University Press, New York.
- Jamison, J., Karlan, D. and Zinman, J.: 2012, Measuring Risk and Time Preferences and Their Connections with Behavior, *Working Paper*.
- URL:** http://www.dartmouth.edu/~jzinman/Papers/Measuring_Risk_and_Time_Prefs.pdf
- 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).
- Keren, G. and Roelofsma, P.: 1995, Immediacy and Certainty in Intertemporal Choice, *Organizational Behavior and Human Decision Processes* **63**(3), 287–297.
- Kovarik, J.: 2009, Giving it Now or Later: Altruism and Discounting, *Economics Letters* **102**(3), 152–154.
- Krawczyk, M. and Le Lec, F.: 2010, 'Give Me a Chance!' An Experiment in Social Decision Under Risk, *Experimental Economics* **13**(4), 500–511.
- 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.
- Loewenstein, G., O'Donoghue, T. and Bhatia, S.: 2015, Modeling the Interplay Between Affect and Deliberation, *Decision* **2**(2), 55–81.
- Machina, M. J.: 1989, Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty, *Journal of Economic Literature* **XXVII**(December), 1622–1668.
- Mongin, P.: 1998, Utility Theory and Ethics, in S. Barberà, P. J. Hammond and C. Seidle (eds), *Handbook of Utility Theory, Volume 1: Principles*, Kluwer Academic Publishers, Boston, pp. 371–481.
- Percoco, M. and Nijkamp, P.: 2009, Estimating Individual Rates of Discount: A Meta-analysis., *Applied Economics Letters* **16**(12), 1235–1239.
- Prelec, D. and Loewenstein, G.: 1991, Decision Making Over Time and Under Uncertainty: A Common Approach, *Management Science* **37**(7), 770–786.

- Rabin, M.: 1993, Incorporating Fairness into Game Theory and Economics, *The American Economic Review* **83**(5), 1281–1302.
- Rotemberg, J.: 2014, Models of Caring, or Acting As If One Cared, About the Welfare of Others, *Annual Review of Economics* **6**(1), 129–154.
- Saito, K.: 2011, Strotz Meets Allais: Diminishing Impatience and the Certainty Effect: Comment, *The American Economic Review* **101**(5), 2271–2275.
- Saito, K.: 2013, Social Preferences Under Risk: Equality of Opportunity versus Equality of Outcome, *American Economic Review* **103**(7), 3084–3101.
- Sen, A. K.: 1961, On Optimising the Rate of Saving, *The Economic Journal* **71**(283), 479–496.
- Sen, A. K.: 1977a, On Weights and Measures: Informational Constraints in Social Welfare Analysis, *Econometrica* **45**(7), 1539–1572.
- Sen, A. K.: 1977b, Rational Fools: A Critique of the Behavioral Foundations of Economic Theory, *Philosophy and Public Affairs* **6**(4), 317–344.
- Sen, A. K.: 2002, *Rationality and Freedom*, 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.
- Tanaka, T., Camerer, C. F. and Nguyen, Q.: 2010, Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam, *The American Economic Review* **100**(1), 557–571.
- 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.
- 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.

- Trautmann, S. T. and Wakker, P. P.: 2010, Process Fairness and Dynamic Consistency, *Economics Letters* **109**(3), 187–189.
- Tversky, A. and Kahneman, D.: 1992, Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty* **5**(4), 297–323.
- Wakker, P. P.: 2010, *Prospect Theory for Risk and Ambiguity*, Cambridge University Press, Cambridge, England.
- Walsh, V.: 1996, Rationality, Allocation, and Reproduction, *Technical report*, Oxford.
- 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.