# A Model of Rational Behavior Combining Processes and Consequences

A Dissertation presented by

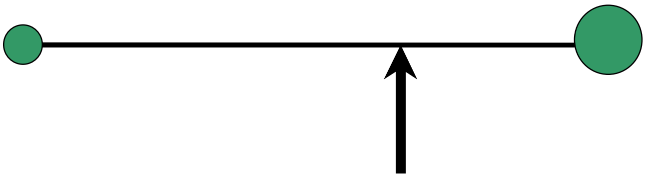
Marc Le Menestrel

to INSEAD faculty
in partial fulfillment of the requirements for the degree of
Ph.D. in Management

June 1999

Dissertation Committee:
Anil Gaba (INSEAD, co-chair)
Ludo Van der Heyden (INSEAD, co-chair)
Christoph Loch (INSEAD)
Raymond Boudon (C.N.R.S., Paris IV Sorbonne)

*To my family and to Swookeun*



*"It would not have occurred to her that an action which is ineffectual thereby becomes meaningless. [...] What mattered were individual relationships, and a completely helpless gesture, an embrace, a tear, a word spoken to a dying man, could have value in itself."*

*George Orwell*

# Acknowledgments

This dissertation is the outcome of a process that is in essence communicational. It was enriched by the sharing of knowledge and emotions with other persons.

I would like to thank H. C. de Bettignies, K. Binmore, A. Boyer, J. Drèze, P. Fishburn, G. Gilbert, K. Goodpaster, J.B. Grize, S. van den Hove, P. Klibanoff, H. Le Blanc, Y. Lengwiller, L. Levy-Garboua, D. Luce, M. Machina, J. March, S. Moss, R. Nau, J. O'Neill, R. Pope, G. Rotillon, R. Sarin, R. Selten, A. Sen, J.C. Thoenig, L. Van Wassehove, P. Wakker, B. Walliser, L. Wathieu, R. Winkler, and P. de Woot de Trixhe for the discussions they kindly accepted to have on these issues. Their comments have been greatly appreciated and led to many improvements of the ideas explored here.

I would like to thank my committee at INSEAD, Anil Gaba, Ludo Van der Heyden and Christoph Loch for their support and guidance. Their continuous efforts and personal involvement have been determinant to the completion of this dissertation.

Raymond Boudon has honored me by accepting to be a member of my committee. His work had long inspired me and it was especially fruitful and encouraging to work with him.

I took as a privilege the opportunity to learn among other Ph.D. students at INSEAD. This community of students is of exceptional value and I would like to thank in particular Canan, Dima and Swookeun, Bjorn, Simon, Johanna, Nicola and Frédéric, Michael and Sumitro as well as former students Miklos, Peter, Alice and Yrjö.

Last but not least, it was a great chance to have my parents, relatives and friends close by. I have been learning from them both intellectually and emotionally and their philosophy has been guiding mine. I also express my gratitude to them. Thanks Sybille for all these instants and what linked them together, for all these *moments*.

# Contents

# Introduction

> *"We do not want to give the misleading impression of attempting here a complete picture of the formation of mathematical models, i.e. of physical theories. It should be remembered that this is a vary varied process with many unexpected phases. An important one is, the disentanglement of concepts: i.e. splitting up something which at superficial inspection seems to be one physical entity into several mathematical notions".*
>
> *John von Neumann & Oskar Morgenstern*

Individuals have ethical concerns: most of us believe lying is wrong.
Individuals have a sense of duty: many believe that democracy is good, and vote.
Individuals have tastes: Ph.D. students like to choose their dissertation topic.
Individuals have emotions: climbers feel the thrill associated to the presence of risk.

In general, ethical concerns, social norms, tastes and emotions participate in motivating behavior. How do they relate with rationality?

In order to explore this question, this dissertation acknowledges that formal models of rational behavior reduce the best interest of individuals to *consequences*. On the other hand, a distinct "dimension" that would *not* be consequential has long been argued to be relevant: "means" are not selected for their "ends" only. Such a "procedural" dimension is of primary importance with regards to ethical concerns, social norms, tastes and emotions. Can we thus construct a formal model that would combine the consequential and the procedural dimensions?

This dissertation proposes such a model. Building on the literature, and in particular on the works of Sen (1995, 1997), Weber (1922), Boudon (1996), and Simon (1955, 1956, 1976), our model considers that *rational* behavior is composed of a *consequential* and a *procedural* dimension. A formal structure is introduced that appropriately reflects this duality. An axiomatic model of rational behavior is then presented and explored.

This "dual" definition of *behavior* combines a *process*—what the individual does—with its *consequence*—what the individual expects to obtain. *Rational behavior* is then defined as the *choice* of the *preferred behavior*. Empirically observable, these preferences are called *behavioral*. At the formal level, the explicit consideration of a process *and* its consequence is identified as a fundamental characteristic of the model. The mere consideration of this dual structure reflects the irreducibility of processes and consequences.

The model is further structured by assuming that individuals have *judgments* for the process of their behavior as well as *judgments* for its consequences. These judgments, distinct from choice, are reflected by *judgmental preferences* on process and on consequences: "process preferences" and "consequential preferences" respectively. When the preferred process leads to the preferred consequence, then such behavior *should* be chosen. When such an *optimal* behavior is not available, the model shows how process preferences and consequential preferences can be revealed *ex post,* i.e. once a rational behavior has been chosen.

This approach allows to define preferences for being an altruist, for not lying, for voting, or for being fair meaningfully and the intuitive assumption that rational individuals choose the preferred behavior can be maintained. It also provides intuitive explanations to empirical "paradoxes" identified with expected utility theory and game theory. These difficulties and paradoxes are argued to reflect the specific influence of *procedural judgments*. Indeed, expected utility theory and game theory formalize rational behavior through preferences over consequences: since they formally exclude procedural judgments, they are purely consequential. This explains why they do not systematically correspond with empirical observation.

Are expected utility theory and game theory therefore useless? Maybe not. I argue that expected utility theory and game theory are of particular relevance for explaining rational behavior. Because they reduce preferences to consequences, they formally "draw the line" between the consequential and the procedural dimensions of rational behavior. A main hypothesis of our approach is thus that expected utility theory and game theory are valid models when properly restricted to the consequential dimension. This validity can be empirically tested and verified only when rational individuals are indifferent towards the process of their behavior. In that sense, these theories need not be always consistent with behavior to be useful formalizations. When individuals have concerns for the process of their behavior, expected utility theory and game theory precisely reveal these judgments which they formally exclude. In order to illustrate these features, the model is applied to two well-known "paradoxes": the "Allais paradox" of expected utility theory, and the "Prisoners' Dilemma" of game theory.

We show that combining expected utility with procedural judgments provides an intuitive explanation to the "Allais paradox", maybe the best-known inconsistency between expected utility theory and empirical observation. According to a purely consequential approach to rationality, the "certainty effect", which is a preference for certainty that is essentially distinct from a diminishing marginal value of, for instance, money, must be interpreted as "irrational". Here, it is modeled as "a process disutility for gambling". The model specifies the extent to which such a process disutility for gambling can be revealed. To this purpose, it maintains the "independence condition between probabilities and consequences" and illustrates how the approach taken is one of extending or "qualifying" expected utility theory, not one of invalidating it.

We show that combining expected utility with procedural judgments provides an intuitive explanation to the well-known discrepancies between game theory and empirical observation in the "Prisoners' Dilemma". According to a purely consequential approach to rationality, "cooperative behaviors" must be interpreted as "irrational". Here, there are shown to reflect the influence of a "social norm" for cooperation. Taking account of process preferences sheds light on how a Prisoners'

Dilemma can be perceived and acted as a “social game” of “common interest” by the individual. Such a “cognitive transformation” can be modeled because process preferences are treated outside the consequence function. This “social game” can be analyzed in a conventional manner, thus providing a new way to combine game theory with ethical concerns, social norms and emotional considerations. When there are several possible equilibria, the Optimal Nash Equilibrium concept refines the notion of Nash Equilibrium to provide a selection criterion. It corresponds to situations where each individual implements his preferred process and reaches the preferred equilibrium. By making a profile of behaviors “focal”, this renders more explicit how a social norm helps individuals to coordinate on the best collective solution for all, even in the context of a single game.

These conceptual explanations are somehow “natural”. This dissertation attempts to explore the “language” that makes them such. The formalism that corresponds to the model is indeed of “peculiar” nature. As already mentioned, the reason is that processes and consequences *depend* on each other through a consequence function. Such a characteristic prevents the independent “realization” of each “component” of behavior. It thus precludes a conventional measurement. For instance, it is not possible to apply the methodology of conjoint measurement developed by the representational theory of measurement. This leaves the structure with a “peculiar” numerical formulation that combines a cardinal scale with an ordinal one. Such a combination remains “open” in a sense that is clarified along the text. These features may be interpreted in the following way. Instead of using formal models to solely capture the consequential dimension of behavior, we should use them to better reveal the singularity of the process of choice, i.e. the specificity pertaining to each individual who performs an act of choice.

At best, these ideas may constitute “an approach toward the understanding” of the influence of ethical concerns, social norms, tastes, and emotions on rational behavior. It may later benefit from other individuals providing their own developments and their own interpretations, so as to render the model less confined by my own ones. I have attempted to introduce as few new notions, notations or semantic usages as possible. A notable exception is the use of the term “process” intended to direct attention to the introduction of “process preferences” or “procedural preferences”. Such a choice of terminology is not at all mandatory and one may prefer to keep using the standard term “action”.

The issue of rational behavior and of its formalization is vast and of relevance in many fields. This raises considerable difficulty in the choice of references and of bibliographic materials. I have tried to select a few but seminal references that have been key to the construction of my argument. This certainly does not give enough credit to many other important and useful pieces. Valuable arguments for the consideration of a “procedural” dimension are numerous and widely spread throughout fields. I am referring to a few I am aware of, although I am not discussing them in all the details they deserve. A more complete review of the literature on this issue is a task I would like to engage in.

The organization of the dissertation is as follows. The first chapter discusses the relevance of combining a consequential dimension with a procedural one, presents the model itself and the illustrations with the paradox of voting and the issue of lying. The second chapter studies the combination of the model with the theory of expected utility proposed by von Neumann and Morgenstern (1944). Implications for the interpretation of formal models are discussed and an illustration is proposed with the utility of gambling and the Allais paradox. The next chapter explores how the model can be combined with game theory. It tries to clarify the normative and descriptive interpretations and provides a refinement of its central concept of equilibrium. It is illustrated by an analysis of the Prisoners' Dilemma. Finally, a brief philosophical discussion concludes.

# Chapter 1
# Combining Processes and Consequences

*“Scientists invent hypotheses that talk of things beyond the reach of observation. The hypotheses are related to observation only by a kind of one-way implication; namely, the events we observe are what a belief in the hypotheses would have led us to expect. These observable consequences of the hypotheses do not, conversely, imply the hypotheses”.*
*W.V.O. Quine*

## 1.1 The need to combine Procedural and Consequential Concerns

> “The standard rationality hypothesis is that behavior can be represented as the maximization of a suitably restricted utility function” (Barbera et al. 1998, p. ix).

As discussed by Amartya Sen (1997, p. 745), this “formulation of maximizing behavior in economics has often paralleled the modeling of maximization in physics and related disciplines”. Indeed, when the formalization of rational behavior became of interest, formal models developed in physical sciences were already much advanced. The analogy between behavior of material objects and behavior of rational individuals became the implicit justification for importing the formalism of maximization initiated in the 17th century.

> “But maximizing behavior differs from non-volitional maximization because of the fundamental relevance of the choice act, which has to be placed in a central position in analyzing maximizing behavior” (Sen 1997, p. 745).

Unlike objects in mechanics, rational individuals are deemed to have intentions and purposes. They have emotions and significant cognitive capabilities. They are influenced by their past and anticipate their future. They are embedded in a social context and are influenced by emotions of others. In some situations, individual rational behavior is not determined by necessity. Individuals neither act at random. They have a free will and their act of choice reflects their freedom. In the words of Sen, they express their *volition*.

How can we place such volitional character “in a central position in analyzing maximizing behavior?” How can formal models, which have been primarily constructed to analyze the regularities of behavior of objects, serve to identify, understand or even explain the contextual dimension of the act of choice?

This requires a better understanding of the notion of rationality and of its formalization. As an attempt in this direction, this dissertation builds on the consideration that formal models of rationality are "too narrow" (Boudon 1996). More precisely, they tend to reduce behavior to consequential considerations[1]. As such, standard hypothesis of rationality exclude considerations for the process of behavior. In a seminal article, Sen (1995, e.g. p. 15) argues for a combination of the procedural and the consequential dimensions while pointing that no formal model of rationality seems to combines them explicitly:

> "The need to combine procedural concerns with those of actual events and outcomes is quite strong".

The research question that underlies this dissertation may then be stated as follows: If there is a distinct procedural dimension that influences the act of choice, how could it be formally combined with the consequential dimension?

If we postulate that there are two distinct dimensions, composed of two types of entities, one needs to refer to them with names. This dissertation refers to *consequences* as the entities of *the consequential dimension* and to *processes* as the entities of *the procedural dimension*. The following discussion also intends to explain and justify this choice.

The opposition between a consequential and a procedural dimension tend to raise passionate debates. For instance, it is often argued that rational behavior *must* be solely consequential. Others argue this is a matter of *semantic*. Some say rational behavior must be *solely* procedural, or else, that it is a matter of *doctrine*.

This dissertation wants to avoid such debates about the reduction of one dimension to the other. It treats the relevance of combining two distinct dimensions as an open research question, subject to a rigorous formulation in an appropriate language.

Propositions have been made to formalize procedural considerations. For instance, we may take a dynamic or evolutionary perspective (see Aumann 1997, for a discussion of the rationality of these approaches). We may formalize procedures themselves, like in models of bounded rationality and of "procedurally rational" individuals (see in particular Rubinstein 1998, Osborne and Rubinstein 1998). Besides identifying and qualifying "appropriate procedures", we may also consider that individuals have concerns for the process by which they reach consequences. In social psychology, approaches have been developed to emphasize the role of processes with regards to fairness (Lind and Tyler 1988). In health-care, Mooney (1994) emphasizes the importance of a "process utility" and Donaldson and Shackley (1997) show how such notion is empirically relevant, although difficult to measure. In cognitive psychology, Gärling et al. (1996) study a travel choice problem and evidence that individuals have concerns for the process of their behavior. In philosophy, Nozick (1993) argues for such a "symbolic" dimension of rationality. Since there are many diverse references, I propose to focus on an approach that makes the distinction be-

---

[1] On the importance of consequentialism for formal models, one may refer to Hammond (1988, 1998), Sarin and Wakker (1994), Shafir and Tversky (1992).

tween *judgments on the consequences of behavior* and *judgments on the process of behavior* while arguing for a combination of the two. It originates in sociology.

## 1.2 Ideal-types of Rationality in Sociology

In *Economy and Society: an outline of interpretive sociology*, Max Weber (1922) conceptually combines the procedural and consequential dimensions of rational behavior by proposing two types of rationality. The first type is called "instrumental-rationality". Behavior is said to be instrumentally rational when it is

> "determined by expectations as to the behavior of objects in the environment and of other human beings" (ibid., p. 24).

Consequences of behavior are identified and alternative courses of behavior are compared according to their consequences. In other words, means are selected according to their ends. However

> "the orientation of action wholly to the rational achievement of ends without relation to fundamental values is, to be sure, essentially only a limiting case" (ibid., p. 26).

More precisely, Weber considers that rational behavior is not consequential when it is

> "determined by a conscious belief in the value for its own sake of some ethical, aesthetic, religious or other form of behavior, independently of its prospect of success" (ibid., p. 24).

Emphasizing the role of individual values, Weber termed this second type of rationality "value-rational". The last part of the definition: "independently of its prospect of success", is explicit about the non-consequential nature of value-rationality.

According to Weber, these two perspectives are "ideal-types" aimed at structuring the explanation of rational behavior. In reality, these types together combine:

> "it would be very unusual to find concrete cases of action, especially of social action, which were oriented *only* in one or another of these ways (ibid., p. 26, italics are his).

If the analysis is reduced to instrumental rationality, and Weber adds: "as formulated in the principle of marginal utility", then "value-rationality is always irrational".

As a result of this analysis, it appears that consequential rationality has long been argued to be a limit-case.

More recently, the integrated approach proposed by Boudon also relies on that distinction. Boudon (1994, 1996, 1997, 1998) argues that models of rational choice are too "narrow", not too perfect:

> "we definitely need a model including a definition of rationality broader than

> the one used in the ‘rational choice model’ if we want to explain social action” (1996, p. 147).

Boudon approaches rationality through the notion of “reasons”: an individual is rational when he has “good reasons to act”[2]. Inspired by Weber, he considers that they are reasons grounded in the consequential dimension and reasons grounded in the procedural dimension. These reasons are of different types, subject to distinct criteria of validity. On one hand, consequential reasons may be true or false. On the other hand, procedural reasons are “beliefs”. They are neither true or false: they *are* by themselves[3].

An example may illustrate this point. Suppose an individual likes more money than less. Then, he “should” prefer €100 to €50 if he prefers more money than less. If he behaves so as to receive €50, then a consequential model concludes he has behaved irrationally. This explanation is “final” and excludes reasons grounded in the procedural dimension.

For instance, the individual may choose €50 because he thereby leaves €50 to another person. If the individual likes to be an altruist, this may be perfectly rational. It merely reveals that the individual has reasons beyond consequential ones. There is no formal reason for willing to be an altruist. In fact, it depends on individuals. It is a matter of “judgment”, whose genesis differs from the formal determination that “100 is more than 50”. Indeed, that an individual likes to be an altruist is a property of himself, not of any substantive entities that can be received as a consequence of behavior.

A distinction thus arises as to the nature of the two types of reasons[4]. If, as it is customary to do in formal models, we reflect reasons by preferences, there may be different types of preferences. Let me recall in that respect a quote from Sen:

> “We need to depart both from the assumption of given preferences (as in traditional choice theory) and from the presumption that people are narrowly self-interested homo economicus” (1995, pp. 17-18).

It results from the analysis of Weber and Boudon that such departure may be limited[5]. In particular, the assumption that preferences for consequences are given may be maintained. We may consider that only preferences for processes, whatever they may be, are contextual and thus not given. It would therefore not be necessary either to depart from the assumption that individuals are self-interested. In one sense, rational individuals may still be assumed to choose what they prefer or, in another sense, rational individuals may still be assumed to always prefer more money

---

[2] On this respect, one may refer to Festinger (1957). See also the approach of Levy-Garboua (1998)

[3] Consequential resons correspond to Type I and Type II while procedural reasons correspond to Type III in Boudon’s model (1996).

[4] Boudon (1996) further discusses this point and provides his own examples.

[5] Other works in sociology are of direct relevance. I think in particular of Granovetter (e.g. 1985). One may consult also Petersen (1992ab).

than less. What would be required is a distinction between choice, preferences for processes and preferences for consequences.

To construct such a model of preferences, we need an underlying structure that would formally reflect these conceptual considerations. To introduce a possible one, the next section builds on the work of Herbert Simon.

## 1.3 Substantive Rationality and Procedural Rationality

Simon (1955, 1956, 1976, 1978, 1979) also distinguishes two types of rationality. The first type characterizes formal models as reflecting a "substantive" approach to rationality:

> "Behavior is substantively rational when it is appropriate to the achievement of given goals within the limits imposed by given conditions and constraints. Notice that, by this definition, the rationality of behavior depends upon the actor in only a single respect - his goals" (1976, pp. 130-131).

Substantive rationality thus addresses the consequential dimension of behavior. It is distinguished from a non-substantive type of rationality that is termed "procedural". Simon defines procedural rationality as:

> "Behavior is procedurally rational when it is the outcome of appropriate deliberation" (ibid.,p. 131).

Deliberation refers to some cognitive and introspective process so as to judge "appropriately" a particular behavior[6]. In this respect, it is important to note that Simon defines "procedurally rational behavior" as an "outcome". While this leads to one of the most prevalent semantic difficulty of the distinction between processes and consequences, it also provides the structure so as how the two dimensions combine. Consider the two sentences:

> A consequence is the outcome of a behavioral process.
> A behavior is the outcome of a cognitive process.

These two types of *processes* must not be confused. The first sentence refers to a "behavioral process" and the second sentence refers to a "cognitive process". These two distinct notions can be closely connected. In this respect, consider the following sentence:

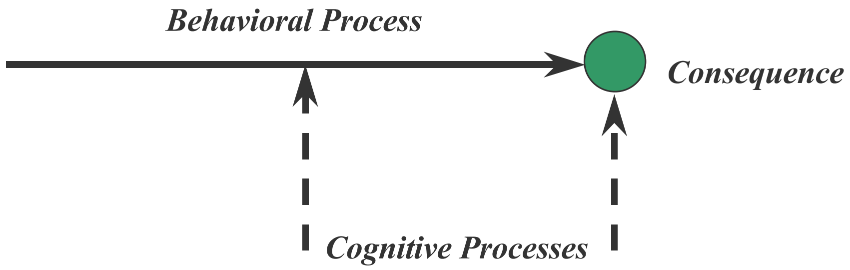
> Implementing a *behavioral process* in order to reach a consequence is the outcome of a cognitive process.

This makes explicit the distinction between the "act of doing" from the "act of thinking". The outcome of an act of doing is a *consequence*. The outcome of an act of thinking is a *judgment*. Procedural rationality may be interpreted as characterizing the judgment on the process of behavior itself, while consequential rationality may

---

[6] Not necessarily among several individuals.

be interpreted as characterizing behavior with respect to its consequence only. This structure is depicted in Figure 1.



1.Behavioral and Cognitive Processes

As such, rational individuals who are concerned by both the process and the consequence of their behavior may be both "consequentially rational" and "procedurally rational" *when they act.* In that interpretation, procedural rationality means that individuals have also considered their judgment on the process of behavior itself. Although ends are the consequences of the means along the behavioral process of choice, these means are ends in themselves along the cognitive process of judgments. This provides an interpretation of the "hierarchy of means and ends" of Simon (1947, p. 73) and illustrates the usefulness of distinguishing the act of doing and the act of thinking.

This reading of Simon differs from most references to his work since, in general, procedural rationality characterizes the procedure by which a choice is made[7]. For that reason, I try not to use the terminology of procedural and consequential rationality. In any case, what is assumed is that individuals have "made up their minds" and that rational behavior, as the "observable" of the model, reflects their judgments.

Although this discussion of the literature leaves aside a detailed discussion about approaches of rationality in the social sciences, it justifies the relevance of combining a consequential and a procedural dimension while introducing the main notions of our model[8].

## 1.4 Presentation of the Model

Our model is presented in five subsections organized as follows. A notion that is later formally defined first appear in italics.

---

[7] One may find in Conlisk (1996) a review of this literature.

[8] March (1988) provides a review of many different perspectives on rationality. Elster (1979) and Frank (1988) are seminal works. Smelse (1992), Dreier (1996), Walliser (1989), Hechter (1995), Vriend (1996), Benett and Howard (1996) constitute a first selection in which many other useful references can be found in different directions. Requier-Desjardins (1995) has an interesting analysis linking with time.

1. *Rational behavior* corresponds to the *choice* of the *preferred behavior*. Preferences among behaviors are called *behavioral preferences*. They are directly observed from *choice*.
2. Each behavior is composed of both a *procedural and a consequential dimension*. Entities of the procedural dimension are called *processes.* Entities of the consequential dimension are called *consequences*.
3. Rational individuals are assumed to have judgments over processes, reflected by *process preferences* (or procedural preferences). Rational individuals are assumed to have judgments over consequences, reflected by *consequential preferences*. These preferences are called *judgmental preferences* to make explicit that they are not directly observed through choice.
4. Whenever such a behavior is *available*, a rational individual always chooses an *optimal behavior,* i.e. a behavior that is composed of a preferred process and a preferred consequence.
5. Preferences over processes can be *revealed* from choice *dependently* on consequential preferences. Similarly, preferences over consequences can be *revealed* from choice *dependently* on process preferences.

These notions are now developed and formalized. I start from behavior since this is the observable of the model.

### 1.4.1 Individual Rational Behavior

By considering individual behavior, I follow an approach known as "methodological individualism". According to methodological individualism, social phenomena must be explained by referring ultimately to individuals' behavior. Individuals are actors who transform the world intentionally and purposively. Social norms may nevertheless be influential. Methodological individualism requires that they are "internalized" by the individual who acts[9].

By considering rational individuals, I am supposing that individuals act according to their preferences. Rational individuals compare available behaviors and choose the one they prefer[10]. Their behavior being assumed to be both intentional and purposive, it is the outcome of appropriate cognitive processes that occurred before the behavior itself. The observation of behavior is expressed as a preference among available behaviors, opening the way for a rigorous formalization of ratio-

---

[9] See on this respect Boudon (1987), Elster (1989ab) and Arrow (1994). I focus here on individuals as physical individuals. Behavior of firms, institutions, or more generally collectivities of individuals, may also be studied as actors through methodological individualism.

[10] *Rational behavior* and *choice* are exchangeable notions in this model.

nal choice[11]. There is no a priori selection of which preferences are true, false, right, wrong, good, bad, etc. for the model to be applied[12].

This allows a first set of notions to be formalized. Each individual chooses among a set $B$ of *available* behaviors. Such choice is reflected by the binary relation $\succsim^B$ which is called *behavioral*. That an individual chooses a behavior $b^*$ over a behavior $b$ is described by the relation $b^* \succsim^B b$. It is assumed that individuals choose among *all* behaviors of the set $B$. Furthermore, it is assumed that the relation $\succsim^B$ is *transitive*. These assumptions can be formulated through the following definitions and axiom. For such general definitions, which are later applied to many different sets, a "generic" notation is used. A generic set is denoted by $S$ and its elements $s, t, u \in S$.

**Definition 1** *A binary relation* $\succsim^S$ *is a* ***weak ordering*** *of* $S$ *if and only if, for all* $s, t, u \in S$ :

***Completeness***: *either* $s \succsim^S t$ *or* $t \succsim^S s$ *or both.*
***Transitivity***: $s \succsim^S t$ *and* $t \succsim^S u$ *imply* $s \succsim^S u$.

It is also convenient to define two other binary relations $\sim^S$ and $\succ^S$ from a weak ordering $\succsim^S$ as follows:

**Definition 2** *For any weak ordering* $\succsim^S$, *for all* $s, t \in S$ :

***Indifference***: $s \sim^S t$ *if and only if* $s \succsim^S t$ *and* $t \succsim^S s$.
***Strict Preferences***: $s \succ^S t$ *if and only if* $s \succsim^S t$ *and* $not(t \succsim^S s)$.

The definition of a weak ordering is applied to behavioral preferences. This constitutes the first axiom of the model.

**Axiom 1** *The binary relation* $\succsim^B$ *is a weak ordering of* $B$.

Individual rational behavior can now be formally defined as the choice of the preferred behavior.

**Definition 3** *A behavior* $b^*$ *is* ***rational*** *if and only if, for all* $b \in B : b^* \succsim^B b$.

At this stage, these assumptions are rigorously standard. The main departure from standard rational choice comes from the structure of the set $B$ of available behaviors. Such a structure is introduced in the following section through the notions of processes, consequences, and the consequence function.

### 1.4.2 Processes, Consequences and the Consequence Function

Buying or selling a product, producing or destroying a good, gambling to win a lottery or not gambling to secure an outcome, cooperating to conduct a collective

---

[11] There are models formalizing rationality where the role of preferences is far from central. These approaches take a more "macro" point of view than the one of methodological individualism. See for instance Nau and Mc Cardle (1991).

[12] Such a selection is required in a purely consequential approach (see e.g. Harsanyi 1978).

action or competing to increase an advantage, investing in a project, voting to elect a representative, climbing a mountain, lying or not lying to earn money, etc., are all examples of individual behavior. Any description of such a behavior involves two dimensions: a transformation of the world—the *process*—and an end state that results from this transformation—the *consequence*.

The process corresponds to what the individual actually does: buying, selling, gambling, voting, cooperating, lying, not lying, climbing, etc.[13] Semantically, a verb usually designates the process. In practice, the process is implemented by the individual himself when he performs the act of choice. It is the dimension which remains under the *control* of the individual. In that sense, the fact that a particular transformation has been implemented necessitates an *act of choice*: if the individual does nothing, nothing happens. The act of choice is thus distinguished from a "decision" to act: one can decide to do something, although he hasn't done it yet, so nobody can know or observe. Indeed, he may later do something else. Here, a decision is not considered in itself i.e., until it is expressed through behavior. When a behavior occurs, something must be carried out, and this is what a *process* is assumed to be[14]. As a result, there is only *one* (behavioral) process per behavior and processes are *singular*.

The end state corresponds to the *consequence* of such process: a product or a good, a lottery, an election, a project, an amount of money, an ascent, etc. Semantically, a noun usually designates a consequence. A consequence is said to be *expected* or *perceived*. While a process refers to a transformation, with some notion of dynamism involved, a consequence has some "substance". An entity is "substantive" when it can be composed with other substance of the same "extensive structure". This characteristic may be contrasted with the "singular" nature of processes[15].

Figure 2 depicts such dual consideration of behavior. A consequence is depicted as a circle in order to provide intuition that it reflects a substantive entity. When the individual himself is also depicted on the figure (for instance as a square distinct from the consequence), the process appears as a *relation* between the individual and the consequence *as a whole*.

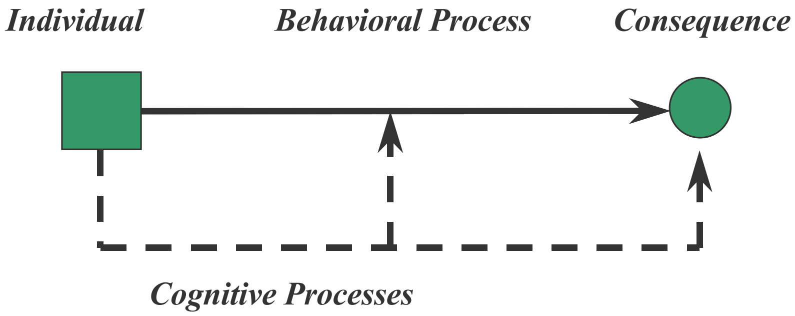
The fundamental departure from standard model has been to consider that behavior is made of these two types of entities. It turns out that the combination of a process and a consequence is not of the same nature than the combination of two "things". Since processes lead to consequences, consequences depend on processes for their realization (Krantz et al. 1971, p. 246). This *dependence* entails a peculiar dual structure, composed of "complementary" or "irreducible" entities. Each entity

---

[13] A process may be called an *action,* as it is customarily done in consequential models, for instance in game theory. The terms *action* and *process* are thus exchangeable. in any case, we have to pay attention not to confuse behavior, process or action and consequences.

[14] Note that this raises a difficulty when an individual defers his behavior and "chooses to do nothing". This issue is related with the notion of the "status-quo" that is discussed briefly in chapter 2.

[15] This is formalized in chapter 2 by introducing the standard operation combining consequences and probabilities. I tend to use "entities" to refer both to "things" or "objects" (as exemples of substantitve entities) and processes or "relations" (as not substantive).



2.Behavior = (Process, Consequence)

is not separable from the other in terms of empirical observation. This has important implications for the assignment of numbers to the entities of *choice*, i.e. for their measurability, since conventional methodology does not apply. This may be better explained by first showing how the condition of *independence of realization of entities* usually applies.

Consider baskets of fruits composed of pears and apples. There are three types of entities: fruits, pears and apples. The *independent* realization of entities means that it is possible to compose three kinds of baskets of fruits: the ones composed only of pears, the ones composed only of apples, and the ones composed of both. Then, by comparing baskets of fruits, we can also compare, for instance, a basket of pears with a basket of apples. This provides a form of commensurability. Although this may be seen very natural, this is not possible in the case of behavior as composed of processes and consequences.

For instance, there is no behavior composed solely of processes that can be compared to a behavior solely composed of consequences. Any behavior *must* be composed of a process and a consequence to be a behavior at all. The trade-off between procedural judgments and consequential judgments occurs cognitively without any substitutability between processes and consequences

Another example, in physical sciences, is the consideration of *moments.* For instance, the consideration of both velocity and position raises known difficulties in terms of measurement (see also Krantz et al. 1971, chapter 6). To overcome these, one usually consider multiple observations so as to work with distribution of probabilities, the observable of the model becoming statistical[16].

We have thus identified the main formal departure of our model in the observation that the *independence of realization of entities* is not fulfilled. The non-substitutability it entails is one manner to interpret the irreducibility of the procedural and consequential dimensions of behavior. I propose to use the term “couple” (of processes and consequences) rather than “pair” to reflect this property.

This dependence between processes and consequences is not arbitrary but derives from the specification of a situation of choice. This is made explicit through a

---

[16] This is later evoked in chapter 2 but it is not studied in this dissertation.

*consequence function* that specifies which process leads to which consequence. The consequence function must thus be taken as given and is specific to the situation under study. Indeed, if any consequence can result from any process, then the intuitive idea of choice must be dropped because any process can lead to any consequence (see Rubinstein 1998, p. 42). The following restates the definition of rationality with processes and consequences.

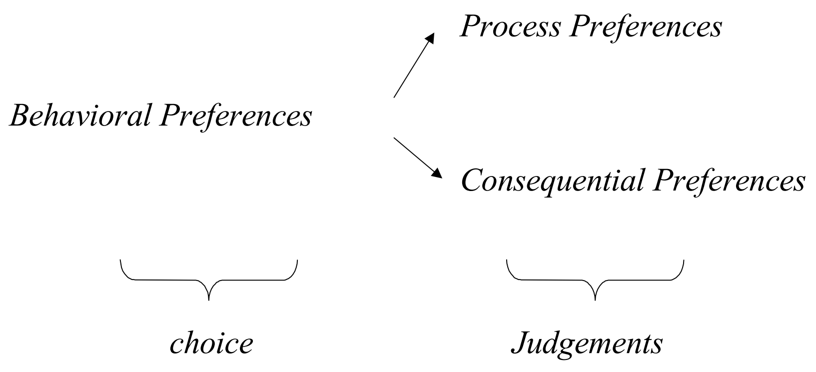
A process is denoted by $a \in A$. A consequence is denoted by $c \in C$. The set $A$ is the set of *processes* and the set $C$ is the set of *consequences*. The dependence between processes and consequences is reflected by a *consequence function* $g$ with domain $A$ and range $C$. A consequence function is written $g : A \to C$, $a \mapsto g(a) \in C$. A *situation of choice* corresponds to the set of *available behaviors* $B = \{(a, g(a)) : a \in A\}$. Individual rational behavior can be formally re-defined as the choice of the preferred behavior, as a couple of a process and its consequence according to a consequence function.

**Definition 4** *A behavior* $(a^*, g(a^*))$ *is* ***rational*** *if and only if, for all* $a \in A$ : $(a^*, g(a^*)) \succsim^B (a, g(a))$.

The model of this dissertation thus does not introduce any new entities apart from the combination of processes and consequences which also appear in standard models. It however introduces distinct notions of preferences by distinguishing behavioral preferences on one hand from consequential and process preferences on the other hand. These latter *judgmental* preferences are the subject of the next subsection.

### 1.4.3 Distinguishing Choice from Judgment: Behavioral and Judgmental Preferences

A fundamental assumption of this model is that *individuals make judgments on both processes and consequences*. In this manner, the model distinguishes the notion of *judgment* from the notion of *choice* (see Figure 3).



3.Distinguishing Choice from Judgements

Judgments reflect an evaluation of processes or of consequences by the individual. As *states of mind*, they are *cognitive*. By opposition, choice is *empirical*. For instance, an individual who dislikes to lie (a judgment on the process) and likes more money than less (a judgment on the consequence), may rationally *chooses* a behavior leading to a lower amount of money. In such a situation, *choice does not correspond with judgments on consequences*. The same reasoning with judgments on consequences can be made. In general, rational behavior—or choice—reflects the combination of both a judgment on processes and a judgment on consequences.

Coming back to the example of altruistic behavior discussed in page 8, the point is that individuals may still do what they prefer. They may even remain "narrowly self-interested" in their consequential judgments. The rationality of an individual who acts in an altruistic manner (i.e. forgoing €50 in the example of page 8) implies that he is preferring to be an altruist, as a procedural judgment. Distinguishing choice from judgments allows an individual to behave in an altruistic manner while still preferring more money than less. Such a distinction occurs meaningfully because of *the dependent realization of processes and consequences* that is proper to the two types of entities of the present model. When there is independent realization of entities, judgments and choice are confused. Think again of the example with baskets of fruits. When a basket is composed only of apples, then [choice] of such a basket is the same as judgment over apples and no precise distinction can be made[17].

The opposition between judgments on processes and judgments on consequences can be understood as an opposition between *right* and *good* (Mongin and d'Aspremont 1998, p. 376)[18]. Judgments about consequences reflect which consequence is better than the other. In that manner, they reflect the values of objects, as entities distinct from the individual. Judgments on processes reflect whether some process is "right or wrong" according to the values of the individual. Their comparison necessitates the reliance of a rule, i.e. "lying is wrong". In this manner, process preferences are of a deontological nature. They reflect a conformity between the process and a rule of behavior, or more generally a social norm that is internalized by the individual. Typical examples are the reluctance to lie, choosing to vote, or simply choosing to leave a tip in a restaurant you know you will never visit again. Anticipating on chapter 2, such a distinction is even more salient when judgments on consequences are "quantitative" while judgments on processes remain "purely qualitative".

If we now turn to the formalization of these notions, judgments are represented by preferences. Procedural judgments are reflected by a binary relation $\succsim^A$ over the set of processes. Consequential judgments are represented by a binary relation $\succsim^C$ over the set of consequences.

---

[17] I put choice within square brackets because it does not correspond to the definition of our model but to a consequential approach that does not take processes into account. In this model, the process of obtaining such a basket should be made explicit and may bear attributes that are influential to real choice. For an example of what such attributes may be, see the discussion in chapter 2 about preferences for intrinsic diversity.

[18] They refer to the consequential dimension as "teleological" and to the procedural dimension as "deontological".

It is assumed that all processes are judged and that all consequences are judged. Besides, preferences are transitive. This is stated in the following two axioms.

**Axiom 2** *The binary relation* $\succsim^A$ *is a weak ordering of A.*

**Axiom 3** *The binary relation* $\succsim^C$ *is a weak ordering of C.*

Before proceeding further, note that standard consequential models only consider consequential preferences. As such, they identify choice with consequential judgments. In that respect, they may define a rational behavior as [choose $a^*$ such that $g(a^*) \succsim g(a)$]. Judgmental preferences must now be linked with behavioral preferences, which is the purpose of the next subsection.

### 1.4.4 Combining Preferences with an Optimality Axiom

The fundamental assumption about the combination of judgmental preferences is that a behavior composed of a procedurally preferred process and a consequentially preferred consequence *should* be behaviorally preferred.

This axiom states as follows:

**Axiom 4** *If* $a \succsim^A a'$ *and* $g(a) \succsim^C g(a')$ *then* $(a, g(a)) \succsim^B (a', g(a'))$.

Such behavior is called *optimal*[19]. Optimality is a *stronger* property than being preferred: a behavior may be chosen while not being optimal. The axiom is a one-way implication that corresponds to a form of "monotonicity" of process preferences and consequential preferences. Because of the dependence of the two underlying entities, its formulation differs from the conventional formulation of "independence of preferences" among entities[20]. Indeed, it does not specifically require process preferences to remained "constant" when consequences are modified.

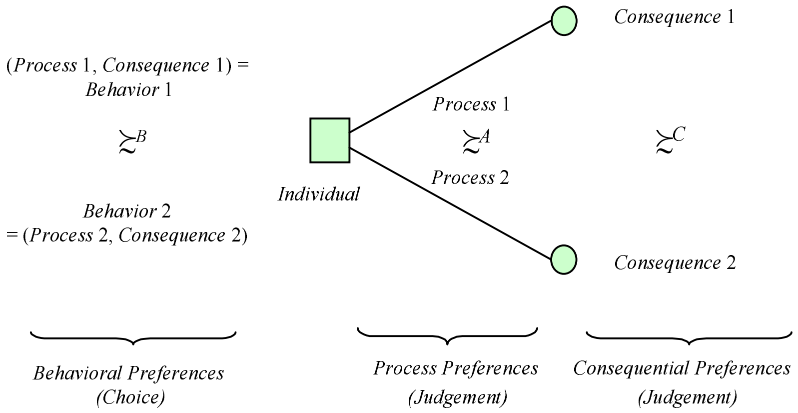
The optimality axiom allow behavioral preferences to be determined by judgmental preferences when process preferences do not "conflict" with consequential preferences (see Figure 4).

When, in situations of *dilemmas,* judgments conflict, the optimality axiom alone does not propose any ex ante solution. This form of indeterminacy may appropriately reflect the volitional character of rational behavior. Nevertheless, the following subsection shows how judgmental preferences can be revealed ex post, i.e., after behavior has occurred.

---

[19] See Sen (1997) for the relevance of introducing a distinction between a maximal behavior (the preferred one) and an optimal one. This property of optimality is possible because the two dimensions are irreducible one to the other.

[20] On the different notions of independence, see for instance Fishburn and Wakker (1995), Keeney and Raiffa (1976), Krantz et al. (1971, chapter 6).



4.Combining Preferences: the Optimal Case

### 1.4.5 Dependent Revelation of Judgmental Preferences

The main idea of the dependent revelation of judgmental preferences consists in taking the contraposition of the optimality axiom. More precisely, this consists in rewriting an expression of the form "$\{a\ and\ c\} \Rightarrow b$" as, for instance, " $\{not(b)\ and\ c\} \Rightarrow not(a)$". In this manner, if $a$ and $c$ are judgmental statements, two such judgments may determine the observable $b$ ex ante. By contraposition, an observable ex post, together with one of such two judgments may also determine the complementary judgment. Formulated as a corollary of the optimality principle, this gives:

**Corollary 1** $\{(a', g(a')) \succ^B (a, g(a))\ and\ g(a) \succsim^C g(a')\} \Longrightarrow a' \succ^A a.$

*Proof: By contraposition of the optimality axiom, we have* $not[(a, g(a)) \succsim^B (a', g(a'))] \Longrightarrow \{not[g(a) \succsim^C g(a')]\ or\ not[a \succsim^A a']\}$. *Since* $not[(a, g(a)) \succsim^B (a', g(a'))] \Longleftrightarrow (a', g(a')) \succ^B (a, g(a))\ and\ not[a \succsim^A a'] \Longleftrightarrow a' \succ^A a$ *, we have* $\{(a', g(a')) \succ^B (a, g(a))\ and\ g(a) \succsim^C g(a')\} \Longrightarrow a' \succ^A a$ □

This corresponds to the *dependent revelation of preferences over processes*. Similarly, *dependent revelation of preferences over consequences* can be formulated as:

**Corollary 2** $\{(a', g(a')) \succ^B (a, g(a))\ and\ a \succsim^A a'\} \Longrightarrow g(a') \succ^C g(a).$

*Proof: Same as above* □

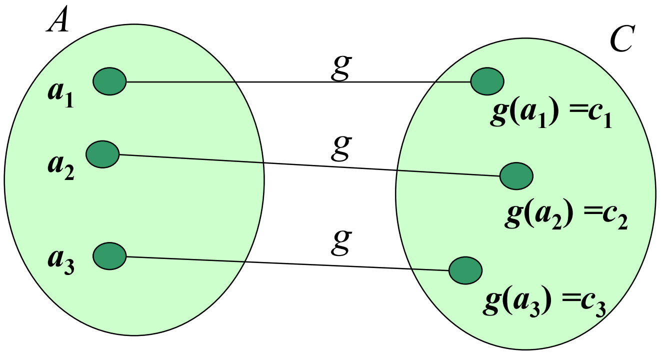
Thus, depending on an hypothesis about consequential preferences, the optimality axiom may reveal process preferences ex post. A behavior that is preferred to another one which possesses a better consequence must possess a preferred process. Similarly, a behavior that is preferred to another one which possesses a better process must possess a preferred consequence. A typical example is to observe an individual refusing to lie and earning a lower amount of money than he could have reached with lying. Such a rational behavior reveals his process preference for not lying. Although each type of judgmental preferences is not separately observable, the structure proposed allows them to be revealed *dependently* on an assumption about the other judgmental preferences.

When consequences are amounts of money, such an assumption merely requires to specify the amounts and to assume that the individual prefers more money than less. In general, it may be more delicate. Chapters 2 and 3 explores how expected utility theory and game theory formally "draw the line" between processes and consequences and thus provide refinements of the structure.

Finally, the corollaries above do not always allow rational behavior to reveal judgmental preferences. As an example, an individual who is observed to lie in order to earn a greater amount of money may indeed have a preference for not lying. What such a behavior reveals is that the "intensity" of his preference for not lying is not sufficient to overcome the "strength" of his preferences for amounts of money. As for its normative counterpart, this form of indeterminacy in the explanation of behavior may conveniently reflect the volitional character of rational behavior.

### 1.4.6 Incomplete Preferences and Conditional Preferences

Another manner to approach the combination of processes and consequences is of interest. The set $B$ of available behaviors differs from the Cartesian product $A \times C$. Indeed, if you consider *a contrario* that $B = A \times C$, then $g$ assigns any consequence to any process and the situation of choice is not specified at all. As a result, we have $B \subset A \times C$ and $\succsim^B$ is complete over $B$ but incomplete over $A \times C$ (see Figure 5).



$\succsim^B$ **is *complete* over $B = \{(a, g(a)) : a \in A\}$**
$\succsim^B$ **is *incomplete* over $A \times C$**

5.Complete and Incomplete Preferences

Sen (1997) indeed represents the volitional character of the act of choice through an incomplete relation over a set of "comprehensive" outcomes. This would be the analog of $\succsim^B$ when considered over the *whole* Cartesian product $A \times C$. However, Sen's approach does not specify the underlying structure as I do. More precisely, it does not identify any *complete* relation over a *determined subset of the whole Cartesian product*, namely the subset engendered by the consequence function itself.

This latter structure indeed allows the corollary of section 1.4.5 to be derived. Because behavioral preferences have been assumed to be *complete* over $B$, the equivalence $not(b \succsim^B b') \iff b' \succ^B b$ holds. If we do not specify such property of $\succsim^B$, then this equivalence does not hold and we cannot obtain the desired result.

Our model also differs from another important avenue initiated by Drèze (1958, 1987) and known as "state-dependence". Applied to the present case, such a methodology would consist at defining consequences *conditionally* on processes. The next step would be to assume some structure for such preferences on conditional consequences. We would then have constructed "process-dependent" preferences in a manner analogous to "state-dependent" preferences. Such an approach requires, however, processes to be reduced to consequences, even in a conditional manner. Our model attempts to take seriously the combination of two entities dependent on each other through a consequence function[21]. It is now proposed to illustrate the model with two examples.

## 1.5 Lying or not lying? A Question of Ethics...

Suppose an individual can earn €100 with lying while he can only earn €50 without lying. Suppose further that no other consequence can derive from his behavior as it is assumed in conventional models of consequential rationality. Then, a consequential approach to rationality requires that the individual lies to be considered rational. Any other behavior is irrational.

The consequential treatment of the problem of lying somehow assumes that individuals have no displeasure for lying per se. However, there may well be argued that an internal need of cognitive consistency within individuals generates such a displeasure, which in turn influences choice. But then, what the individual should do in such an ethical dilemma?

The prescription of Kant's moral philosophy is that there is no right to lie in principle nor in practice. Ethical duties are "categorical imperatives" and state that individuals should not act according to a principle if they do not desire that everybody follows that principle. Indeed, if everybody was lying, language would not mean anything. For Kant, this proves that an individual has no right to lie. Therefore, an individual cannot want to lie, and shall therefore not lie, even if he reaches a worse consequence.

Distinct from the purely consequential approach as well as from the purely procedural approach that propose one rational solution that lacks any intuition, our

---

[21] The study of partial orders has been pointed by von Neumann and Morgenstern in the *Theory of Games* as a possible extension to their concept of utility (see in particular §65.5). One may consult also Skiadas (1997) and his references. On state-dependence, see also Fishburn (1973); Kadane and Winkler (1988); Schervish et al. (1990). About preferences on subsets of product sets, see Sainfort and Deichtmann (1996) and their references.

model does not offer any solution *a priori* to such dilemmas but is so intuitive that it seems simple common sense. It nevertheless allows to express these situations in terms of preferences and this may be of interest to go further. In chapter 3, a game with process preference for cooperation is studied. These preferences may be preferences for not lying that result from pre-play communication. This again open further opportunities to model rational communication.

## 1.6 The “Paradox” of Voting

Consider the question: why does an individual vote?

The common sense is to answer that “individuals vote in order to elect their representatives”. The answer is consequential and is perfectly correct. However, it is a collective answer and as such, does not respect methodological individualism. Individuals do not vote together, they choose to vote individually. The “paradox” appears only when we maintain both consequentialism and methodological individualism. If we refer to a specific individual in a particular context and ask why *this* individual votes, we can no longer answer: “*this* individual votes to elect a representative”, for his influence of the outcome of the vote is almost null[22]. Any small incentive not to vote (like a planned weekend with friends) overcomes the interest of such a consequence. Consequential reasoning thus concludes that it is not rational to go voting and thus, that an individual who goes voting is irrational.

The difficulty to satisfactorily explain social behavior with consequential models of rationality is exemplified in this paradox of voting. On one hand, our societies evolved through the development of democracy as the expression of individuals’ choice for their representatives. On the other hand, standard models of rational behavior are not capable of formulating the rational character of voting behavior. For these reasons, the issue has been called a paradox.

In order to avoid the difficulty, we could discuss the “almost null” and consider the difference between no consequence and almost none. We could also restrict ourselves to collective answers, considering somehow that applying methodological individualism to such a collective issue has no meaning. We could also answer that there is no reason for an individual to vote, and that such an act is deemed to be “irrational”. We could simply dismiss the question. Here, we consider that rational behavior is not reduced to the consequential dimension.

An individual is rational if he does what he prefers. If he prefers the pleasure of a week-end with friends, then the fact that he goes voting in spite of such a pleasure reveals that he has some preference for the process of voting itself. Whether he is right or wrong to have such preference is another question. Whether such preference is a pleasure to vote or a displeasure not to vote is irrelevant here. In any case, that such a preference is based on a belief and not on the quantitative evaluation of consequences does not make it an “irrational preference”. Individuals’ process pref-

---

[22] The probability that his single vote is influential is rapidly vanishing with the numbers of voters.

erences drive them to go voting, because they want to and prefer to, even if they know that the consequence is “almost null” or unfavorable in a particular context[23].

Naturally, the question “why do individuals have these process preferences?” remains. Such a question is about the emergence of judgments as well as about their justification, not about the explanation of behavior. This model formalizes the latter, not the former.

## 1.7 Additional Remarks on the “Peculiarities” of the Model

I would like to draw attention to the nature of the arguments supported by this model. As it may have appeared to the reader through the examples, they have a counterfactual construction. Without entering into a detailed analysis, a counterfactual argument builds on the consideration that a fact did not occur so as to infer some information about the world transformed. In the example introduced in page 8, the reasoning is as follows. If the individual had not been an altruist, then he would have chosen to receive €100. Since he has chosen to receive €50, then he must be an altruist. This indicates the “peculiarity” of the logic that underlies this model.

A special mention should also be made to the type of predictions supported here. Suppose in the situation described above that the individual is an altruist. The model cannot predict whether he will choose to receive €100 or €50 (no available behavior is optimal). However, if the individual does the latter, then it is possible to conclude that such a hypothesis was right. Suppose now that the individual is not an altruist. Then, the model predicts the individual will behave as to receive €100. However, if he does so, it does not confirm the hypothesis, since he may have been an altruist, but “not enough”. As a result, it appears that an hypothesis on process preferences that allows a prediction cannot be confirmed through experience, while an hypothesis that is not sufficient to form a prediction may be confirmed ex post[24].

## 1.8 Conclusion to Chapter 1

In this first chapter, I have tried to situate models of rational behavior in their context. Building on the literature, I justified the relevance of an approach to rational behavior that combines a procedural dimension with a consequential one. Exploring the integration of procedural judgments within an appropriate formal model has

---

[23] I do not claim to be the first to argue that voting is rational because individuals may prefer to vote as an ethical value (see for instance Boudon, 1997). See also Brennan and Hamlin (1996) and (Sen, 1997).

[24] Kahneman and Miller (1986) use counterfactuals to study norms. A counterfactual reasoning also defines an “appropriate” or “satisficing” solution in Simon (1976, p. 140). On the link with ethical values, see for instance Sen (1986). For the seminal treatment in intentional logic, see Lewis (1973). Both points deserve further investigation.

been identified as the objective of this dissertation. This first chapter presents such a model.

The approach is based on a modification of the "entity of choice". It considers that behavior does not reveal choice among "objects", but choice among dual entities composed of a process and of its consequence according to a consequence function. Instead of considering that rational individual behavior can be reflected by preferences over consequences only, rational individual behavior is reflected by preferences over processes combined with preferences over consequences.

All notions to the exception of the distinctions among preferences are already present in consequential models. However, the structural modifications it entails have significant implications.

It allows a distinction between choice and judgments to be explicit. It thus distinguishes among different notions usually subsumed under the same formal concept of *preferences*. This is further refined by distinguishing judgments on processes from judgments on consequences. As such, judgments on the ends are distinguished from judgments about the means themselves. In other words, the means are not judged only for the end they serve, but also for themselves. This makes explicit the ethical maxim that the ends may not justify the means.

Formally, the mere consideration of a preference relation over dual entities of the form (process, consequence) significantly modifies the conventional structure of formal models of rational behavior. Because the procedural and the consequential dimensions of behavior are dependent on each other, there is a form of irreducibility between processes and consequences. The trade-off between them cannot be directly observed as it may be among, for instance, pears and apples. Axiomatically, these judgmental preferences combine to form behavioral preferences in a one-way implication. In that manner, the model remains "open".

Along this first chapter, many questions have been deferred. Among the most important ones:

- Why can't process preferences be integrated within consequential preferences?
- Are expected utility theory and game theory necessarily consequential?
- How may consequential preferences be quantitative on a continuum while processes would remain qualitative and discrete (such as for instance lying or not lying, voting or not voting, etc.)?

In the next chapter, the consequence function is probabilistic and as such, allows consequential preferences to be measured quantitatively. This provides clues towards answering these questions.

# Chapter 2
# Processes and Probabilistic Consequences

> *"Thus far, we have been discussing the properties and qualities of things mainly in so far as they may be abstracted from the process in which these things are always changing their properties and qualities and becoming other things".*
> *David Bohm*

## 2.1 The Independence Condition between Probabilities and Consequences

The approach to utility proposed by von Neumann and Morgenstern in the *Theory of Games and Economic Behavior* (1944) has had an immense impact on formal models of rational behavior[25]. By introducing probabilities in their axiomatization, their model allow preferences to be measured as *quantitative* properties of the entities of choice, i.e. to measure preferences on an interval—also said cardinal—scale. Nowadays, the general principle of expected utility theory has been extensively studied and developed. The general view is that it remains the proper *normative* model of rational behavior.

A central idea of the theory is the "*independence condition between probabilities and consequences*". This condition assumes that the utility of consequences is independent from probabilities and thus requires a form of "separability" of consequential preferences from probabilities. The axioms of expected utility theory can be formulated such that this independence condition together with a weak ordering assumption (see chapter 1) and a continuity property, ensure the *quantitative measurement* of utility. In the expression of the theory when probabilities are subjective—the theory of *subjective* expected utility—the independence condition takes the form of the "sure-thing principle". It is central in Savage (1954), which rejoins early works of Ramsey (1931) and de Finetti (1937). It also remains at the heart of the later formulation of Anscombe and Aumann (1963). This dissertation does not further discuss models with "subjective probabilities"[26].

In their analysis of the notion and of its development, Fishburn and Wakker (1995) explain how and why

"independence, in all its simplicity and subtlety, has been the most central idea

---

[25] Quotes are referring to the third edition, dated 1953. An insightful retrospective is given by Fishburn (1989). Most formal developments can be found in Fishburn (1982). A review is also Fishburn (1988b). An exposition of the principles is Kreps (1988).

[26] A review can be found in Fishburn (1981). A review of non-expected utility models can be found in Fishburn (1988a). More general, another review is Fishburn (1989b).

in risk theory since mid-century".

Given some preferences, expected utility theory allows other preferences to be formally deduced because they follow a structure and system of axioms that is "closed". This is appropriately called a criterion of *consistency* of preferences. As a result, a rigorous empirical investigation of the correspondence between the theory and actual rational behavior is possible even, for instance, when the diminishing marginal value for money of a particular individual is not known.

Indeed, systematic and reproducible empirical violations of the theory have been rapidly identified. Although expected utility theory is *normatively* appealing, it is however failing *descriptively*[27].

Most attempts to reconcile expected utility theory with empirical evidence have proposed to *relax the independence condition between consequences and probabilities* that lies at the heart of model. Unfortunately, these "*non-expected utility*" theories loose most of the normative power of the original theory. They indeed deviate from the fundamental structure of *linearity* that provides expected utility theory with its mathematical elegance and high tractability[28]. Such structure is arguably part of the intuitive appeal of expected utility theory and of its axioms[29].

This leads two of the most prominent researchers in the field to argue that "the normative and the descriptive analyses cannot be reconciled" (Tversky and Kahneman 1988, p. 167). It then became an important subject of research both formally and empirically: "the tension between normative and descriptive considerations characterizes much of the study of judgment and choice" (Kahneman and Tversky 1984, p. 341)[30].

Although expected utility theory formalizes the *consistency* of rational behavior, individuals are not consistent with themselves. They adapt and evolve. Their behavior depends on the context, on their experience, on their knowledge of what they like and of what may make them happy. Individuals behave depending on the way they perceive and expect things. Besides, different individuals are not consistent with each others, even in the same context. How could a normative theory of rationality be descriptively valid for everybody and every context? For instance, some individuals violate expected utility theory while others don't in the Allais paradox. A theory that would recommend only one rational behavior in such a situation could not be descriptively valid for all individuals. Adaptive behavior, and thus, *contextual* specificity, seems descriptively unavoidable.

---

[27] The most-weel-known violations being pointed by Allais (1953) and Ellsberg (1961). About this vast subject, see for instance Schoemaker (1982); Bell et al. (1989); Luce (1992); Luce and Winterfeld (1994); Camerer and Weber (1993). Beard and Beil (1994) is specially relevant in the present context.

[28] This mathematical notion of linearity should not be confused with the notion of linearity of a function on a real variable (such as money). A formal definition is provided later on.

[29] See for instance Fishburn (1988a) for a review as well as Machina (1987ab).

[30] See also for instance Tversky and Kahneman (1974); (1981); Kahneman and Tversky (1979). A recent account of the role of A. Tversky in the ascent of "behavioral economics" is Laibson and Zeckhauser (1998).

Such a contextual approach to rational behavior should not overlook the usefulness and legitimacy of an approach based on *consistency*. Indeed, without any consistency at all, what could be said about rational behavior? What could be explained and expected?

Attempting to integrate some adaptive dimension while avoiding any complete relativism, the idea is to combine consistency and context specificity in a model of rational behavior. It is proposed that judgments on consequences are *consistent* and do not depend on the context, while judgments on processes may change across contexts. Judgments on consequences may be interpreted as reflecting properties of the consequences themselves. They "remain the same" when consequences "remain the same". Formally, they follow the structure and axioms of expected utility theory. On the other hand, judgments on processes may be interpreted as reflecting the properties of individuals themselves. They change when individuals change. As in chapter 1, process preferences and consequential preferences combine in behavior. This model is proposed as a tool to make rigorous the identification of the contextual specificity of behavior.

The next section introduces the notions supporting this line of argument. It is then developed and formalized in a following section, which also include interpretive comments. Finally, a section illustrates the model with the (dis)utility of gambling and the Allais paradox.

## 2.2 The Role of Probabilities in Expected Utility Theory

The section recalls the role of probabilities in expected utility theory as allowing an interval measurement of consequential preferences. It also shows the limits of such a methodology so as to introduce the relevance of combining expected utility theory with a procedural dimension.

### 2.2.1 Different types of scales

From 1932 to 1940, a committee of the British Association for the Advancement of science debated the problem of measurement. In particular, they studied the question "is it possible to measure human sensation?" In a seminal article, Stevens (1946) did much to clarify and synthesize the matter by proposing a definition and a classification of scales of measurement[31].

According to Stevens, measurement refers to the "the assignment of numerals to things so as to represent facts and conventions about them" (ibid, p. 680). There are many different ways to assign numerals to entities, leading to different scales:

"[These scales are] possible in the first place only because there is a certain

---

[31] I do not discuss here the evolution of the approach since Stevens. Latest developments are referred to in Luce (1996). A detailed mathematical review is Luce and Narens (1985), and a more concise one is Luce and Narens (1987). One may also refer to Krantz et al. (1971). Schwager (1991) provides an assessment of the representational theory of measurement.

> isomorphism between what we can do with the aspects of objects and the properties of the numeral series. [...] The isomorphism between these properties of the numeral series and certain empirical operations which we perform with objects permits the use of the series as a *model* to represent aspects of the empirical world" (ibid, p. 677).

Stevens identifies four types of scale: *nominal*, *ordinal*, *interval*, and *ratio* scales. In terms of semantic, "both ordinal and interval scales have at times been called intensive, and both interval and ratio scales have sometimes been labeled extensive" (ibid, p. 678). I restrict here the use of an *intensive* measurement for ordinal scales. Following Luce (e.g. 1996), I later use the word "extensive structure" for interval scales or ratio scales.

*Nominal* scales are merely "labels" that distinguish things, or more generally speaking, "entities"[32]. For these, "words or letters would serve as well". Therefore, "this scale form remains invariant under the general substitution or permutation group (sometimes called the symmetric group of transformations)" (Stevens 1946, p. 678).

Introducing the notion of order, *ordinal* scales reflect a *ranking* of entities. As an example, if a pear has a better taste than an apple, it is assigned the rank 1 while the apple is assigned the rank 2. "Since any 'order-preserving' transformation will leave the scale form invariant, this scale has the structure of what may be called the isotonic or order-preserving group" (ibid, p. 679). The next sub-section presents how early models of rational behavior rely on such ordinal scales.

The type of scale determines which statistics are meaningful. Most of these statistics are not applicable if the underlying measurement is ordinal (ibid., p. 679):

> "in the strictest property the ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the relative rank-order of data. [...] When only the rank-order data is known, we should proceed cautiously with our statistics, and especially with the conclusion we draw from them."

Providing more structure, *interval scales* attempt to reflect *the extent to which entities differ* (ibid., p. 679):

> "with the interval scale we come to a form that is 'quantitative' in the ordinary sense of the word. Almost all the usual statistical measures are applicable here, unless they are the kinds that imply a knowledge of a 'true' zero point."

Corresponding to the general linear group of transformations, interval scales provide a mathematical structure that is particularly convenient in term of tractability. They however require "a determination of equality of intervals or differences" (ibid, p. 678). For models of rational behavior, the issue is to find a proper empir-

---

[32] Semantic is relevant when one wants to distinguish entities that are "relational" from entities that are "substantitve", or equivalently "extensive". The terms "things" or "objects" tend to drive intuition to a substantitve interpretation.

ical operation that would allow comparison of the type: "I prefer pears to apples 'more' than bananas to oranges" to be empirically meaningful. The distinct solutions to this issue, proposed by Pareto and later by von Neumann and Morgenstern is later presented.

I shall also mention that ratio scales again refine the interval scale by also reflecting comparisons of *ratios* of the numbers that are assigned to entities. As such, *ratio scales* are distinguished from interval scales by the specification of a natural zero. They are the most used in physics and may be confused with interval scales. Stevens evokes *dates* as an example of an interval scale while *periods of time* form a ratio scale. Indeed, differences between dates are compared while their ratio does not mean anything. On the other hand, we can easily have a period of time twice another. For rational behavior, there are difficulties for a natural zero to be formally defined, although it is empirically demonstrated that individuals use one[33].

### 2.2.2 The Solution Proposed by Pareto: Comparing Preference Differences

The classical solution for constructing an interval scale in the context of individual preferences has been proposed by Pareto (1906). It assumes that individuals are directly comparing the extent to which they prefer something to something else. For instance, an individual states "I prefer pears to apples 'more' than bananas to oranges". Denoting such comparison $\gg$, the Pareto solution introduces the basic operation:

$$(pear, apple) \gg (banana, orange). \tag{2.1}$$

If we consider a scale $u$ that assigns a number to fruits, such a scale must preserve these comparisons of differences. This can be written:

$$\begin{array}{rcl} (pear, apple) & \gg & (banana, orange) \\ & \Leftrightarrow & \\ u(pear) - u(apple) & > & u(banana) - u(orange). \end{array} \tag{2.2}$$

Through a direct comparison of differences in preferences, Pareto thus constructs a utility functional that is measuring preferences on an interval scale. Without entering into the technical details of this solution, the issue of the link with observation is raised in the following sub-section[34].

---

[33] The issue is slightly dealt with later on.

[34] Fishburn (1970, chapter 6) provides all necessary technical details on this point, see also Fishburn (1989). For a discussion of the link with a functional that is additive in the different kinds of fruits, one may refer to Fishburn and Wakker (1995).

### 2.2.3 The von Neumann and Morgenstern' Solution: Comparing Probabilistic Consequences

The difficulty of the solution proposed by Pareto resides in the empirical nature of the comparison expressed in expression (2.1). How can such a judgment be obtained? Following Stigler (1950, p. 381):

> "we can construct a unique total utility function if the consumer can tell us the magnitude of the utility gained from one indifference curve (I1) to a second (I2) relative to the utility gained by a move from (I2) to (I3)."

When Stigler speaks about a "utility function", he means the scale $u$ above, that is a functional that assigns a number to each entity on which preferences apply. Stigler perfectly illustrates that the Pareto solution is not based on *choice* as the observable primitive of the model. Individuals are assumed to "tell" what they prefer, they are not required to *act* according to such preferences.

For von Neumann and Morgenstern, this departs from the principle that rational behavior is the only observable on which the formal model should rely. In itself, a verbal statement does not constitute an observation of differences in preferences[35].

Von Neumann and Morgenstern thus reject the Pareto solution (von Neumann and Morgenstern 1953, pp. 23-24):

> "since it does not seem that this relation [an equality relation for utility differences] is really a 'natural' one—i.e. one which can be interpreted by reproducible observations—the suggestion [of Pareto] does not achieve the purpose".

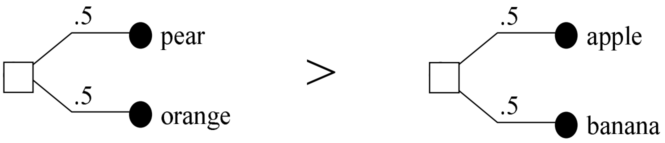
It is indeed the purpose of von Neumann and Morgenstern to introduce a relation of preference that considers choice the legitimate empirical observable.

Von Neumann and Morgenstern start from *preferences over probabilistic consequences*. These probabilistic consequences consist of combinations of consequences with probabilities[36]. As an example consider the following comparison between a pear and an orange obtained with probability .5 each, and an apple and a banana obtained with probability .5 each:

---

[35] Von Neumann's position about the axiomatic method appears clearly in the following statement (von Neumann, 1951, p. 2): "Axiomatizing the behavior of elements means this: We assume that the elements have certain well-defined, outside, functional characteristics; that is, they are to be treated as 'black boxes'. They are viewed as automatisms, the inner structure of which need not be disclosed, but which are assumed to react to certain unambiguously defined stimuli, by certain unambiguously defined responses".

[36] Von Neumann and Morgenstern refers to "utilities". I intend to use instead "probabilistic consequences" in this exact sense in order to avoid confusion between "utilities" in the von Neumann and Morgenstern sense and a "utility function" in the conventional sense.



Assuming that such a preference was observable, von Neumann and Morgenstern show that we can then construct a scale $u$ according to which the above preference is written:

$$\frac{1}{2}u(pear) + \frac{1}{2}u(orange) > \frac{1}{2}u(apple) + \frac{1}{2}u(banana). \tag{2.3}$$

*Algebraically*, this inequality is then equivalent to

$$u(pear) - u(apple) > u(banana) - u(orange). \tag{2.4}$$

As a result, the consideration of preferences over *probabilistic consequences* restricts the measurement of preferences to an interval scale.

Differences in preferences are not assumed as an empirical primitive as they are in the Pareto solution. Differences in preferences are derived *algebraically* from consistent preferences over probabilistic consequences. These latter preferences are assumed to be empirically observable by von Neumann and Morgenstern. Because it is more “naturally” linked with observation, this procedure is more legitimate than the one proposed by Pareto. Nevertheless, the somehow “subtle” distinction between direct comparison of preferences differences and their algebraic derivation from simple comparison between probabilistic consequence has created a considerable amount of confusion[37].

Like pears and apples in the above example, qualitatively distinct entities can be assigned a quantity that represent both preferences and differences in preferences. This is because they are subsumed under the same type of entities through the combination with probabilities, and that these *probabilistic consequences* form a homogeneous continuum.

The notion of a *homogeneous continuum* is central in the issue of measurement[38]. It consists in an structure, called extensive, that is composed of entities that are as “small as you wish” and that are “indistinguishable” from each other. A basic principle of quantitative measurement is then to compare entities of such a structure by their “quantity”. Intuitively, one may see the procedure as constructing a one-to-one correspondence between entities across disjoint subsets of the structure. Then,

---

[37] In several occasions, Fishburn clarified the matter and this section heavily draws on his work. For the original argument, see for instance Fishburn (1970; 1989), Fishburn and Wakker (1995). On may also consult Malinvaud (1952). These issues are further discussed in the next section.

[38] See for instance Luce and Narens (1987) and the references already provided.

for each two subsets, one may contain entities not related to the other, this providing some criterion that this subset is "bigger" than the other. Once the structure is formulated in terms of disjoint sets, and that the reasoning above is applied to any two disjoint sets of the structure, it becomes possible to construct an isomorphism between the structure and the set of real numbers[39].

The consideration of an operation among entities that always result in another entity of the extensive structure ensures that entities are indistinguishable[40]. This is precisely what von Neumann and Morgenstern propose through the combination of consequences with probabilities (this is their second "natural operation"). In this manner, the indistinguishable entities of the extensive structure of von Neumann and Morgenstern are "shifts" in probabilities. A so-called "Archimedean axiom" ensures that these "shifts" can be as "small as you wish".

In that light, the notion of "substantive" entities and of "extensive" entities are reflecting the same idea. It consists in an assumption that the "composition" of these entities do not modify their intrinsic properties. The following two points are then proposed for consideration:

1. nothing is said about whether it is "better to have more" or "better to have less" of these entities.
2. it is suppose that no distinction is made between the properties of a "whole" and the properties of its "parts"[41].

I now propose to illustrate how difficulties encountered with expected utility theory can be interpreted in light of these two points, i.e. that these "abstracted" properties may be relevant in terms of preferences.

### 2.2.4 Judgments Abstracted by the von Neumann and Morgenstern Solution

The first type of judgments that must be excluded from the methodology briefly discussed above is between a positive and a negative extensive structure. Intuitively, a

---

[39] This corresponds to a fundamental result in mathematics known as Hölder's theorem. Luce (1996, p. 79) presents it as follows: "every Archimedean ordered group is isomorphic to a subgroup of the additive reals (or equally, of the multiplicative positive reals). Such structures, called extensive, consist of a set $A$ of elements that exhibit the attribute to be measured; an ordering of $A$, $\succsim$, by that attribute— so if $a, b \in A, a \succsim b$ means that $a$ exhibits at least as much of the attribute as does $b$; and a binary operation $\circ$ on $A$ of combining elements so that $a \circ b$ exhibits the attribute. Mass and length are prototypical". Krantz et al. (1971) is also a seminal reference.

[40] One may refer to Luce and Narens (1987, p. 1531): "For each pair of objects, $a$ and $b$, in the qualitative domain, there exists an automorphism of that domain that takes $a$ into $b$. This proposition is the characterization of homogeneity used in mathematical logic, and it can be shown that for particularly powerful languages describing the domain, it is equivalent to saying that the objects of the domain are indistinguishable from one another."

[41] The first point builds on the notion of a "positive" or "negative" extensive structure (see for instance Krantz et al., 1971). The second on the issue of extensive versus intentional definitions. There is no substantive novelty here (e.g. von Neumann and Morgenstern, 1953, §67).

positive structure is made of a homogeneous substance for which the “more the better” while a negative structure is made of a homogeneous substance for which “the less the better”. An homogeneous continuum must be interpreted either as positive, either as negative. Indeed, if it is both, then it is not homogeneous but has a singularity which corresponds to some “natural zero”. This restricts the formulation of a problem. For instance, it is not possible to integrate the distinction between gains (the more the better) and losses (the less the better) in the measurement of preferences. As a result, entities to be judged must be interpreted in terms of some “final wealth” whose increase is a gain and whose decrease is a loss. As there is no distinction between gains and losses in expected utility theory, there is no notion of status-quo either, while empirically, it is shown to play a significant role[42].

This is of importance where fairness judgments are concerned. If a process if fair and another is unfair, then the “more fair, the better” while “the less unfair the better”. Considering such a dichotomy—typical for ethical judgments, creates a singularity that rules out the homogeneity of the continuum, and thus the expected utility theory.

Turning to the second point related to the whole and its parts, the following quote from Stigler is enlightening, although it should be interpreted cautiously because it does not refer to probabilistic consequence:

> “When it was suggested that the marginal utility of the last yard of carpet necessary to cover a floor was greater than that of fewer yards, the theory was modified to make the covering of the entire floor the unit of utility analysis” (1950, p. 395).

The above quote illustrates that the methodology of measurement does not allow to make a distinction between the whole and its parts. In the context of expected utility theory, this issue takes the form of the “certainty effect” or of the “potential effect”. The “certainty effect” reflects that a shift between a probability .8 and probability .9 is “distinguishable” from a shift between probability .9 and probability 1. This last example, related to the (dis)utility of gambling and to the Allais paradox and its “certainty effect” is further developed as an application in the last section of this chapter.

Finally, another example can be constructed with basket of fruits. When it is assumed that pears, apples, and combinations of pears and apples can all be subsumed under the notion of “baskets of fruits”, and that an operation of substitution is deemed to occur between pears and apples, then any *intrinsic preference for diversity* must be abstracted. Again, we have a similar form of judgment that must be excluded from expected utility theory.

My argument is that these abstracted judgments do not imply to relax the independence condition between probabilities and consequences.

---

[42] See for instance Tversky and Shafir (1992), Luce and Winterfeld (1994, pp. 270, 276), Luce (1996, p. 84). The issue is evoked in Allais (1953), Savage (1954), Kahneman and Tversky (1979).

### 2.2.5 Preserving the Independence Condition between Probabilities and Consequences

That expected utility theory excludes the distinction between “the more the better” and “the less the better” as well as excludes any qualitative distinction between the whole and its parts is traditionally interpreted as violations of the independence condition between probabilities and consequences. Typically, a shift of probability from .9 to 1 induces a higher value for the consequence at stake. A fair process induces a higher value for the outcome. A preference for diversity induces a higher utility for the last substituted apple, the utility of the last piece of carpet is higher than the one of the pieces before, etc.

There is another way to consider these issues. A yard of carpet is a yard of carpet and may have a utility in itself, i.e. as a property of the entity “a yard of carpet”. If any two yards of carpet are exchanged one for the other, then it surely does not matter. Whether a yard of carpet finishes (or does not finish!) to cover the floor, it is matter of the context in which the behavior “placing a piece of carpet” occurs. It is not a matter of “this piece of carpet”. I argue that it may be reasonable to exclude such property (finishing to cover the floor) from any yard of carpet and thus to maintain the homogeneity of the primitive structure.

A similar argument can be made in the context of probabilistic consequences. A sum of €1M may be assumed to remain a sum of €1M, independently of how it is acquired. Indeed, any unit of €1M could be exchanged with another one without affecting the situation at all. Therefore, it may not be that the judgment on €1M changes with probabilities[43].

When pears are substituted for apples in a basket of fruits, the assumption that the last pear removed is not distinguishable in itself from the previous ones seems relevant. The presence of a preference for diversity, that is somehow “triggered” by the removal of the last pear, has nothing to do with the properties of pears and apples. It depends on the individual himself and on the context in which he is judging things. This again argues for maintaining the (abstract) homogeneity of the system of utilities.

Of course, a finished floor, a diversified basket of fruits or a sure outcome do matter. But these properties that are excluded from the measurement because of the mere consideration of the primitive extensive structure as an homogeneous continuum cannot be re-integrated without major logical difficulties.

How can we model these properties then?

They seem to be properties of individuals themselves but the introduction of individuals as substantive entities within the formal model would not solve the case.

---

[43] This may be usefully related to Simth’s distinction between “value of exchange” and “value of use”. Stigler (1950) provides an historical account about how the willingness to measure utility has left aside such a distinction. The form of exchangeability evoked (it is not formally developed here) corresponds to de Finetti (1974). There always remain the solution of transforming probabilities themselves, which is one way to look at non-expected utility approaches.

The following proposes the express these properties as properties of the *relation* between the individual and the consequences.

### 2.2.6 Relating Probabilistic Consequences to Individuals through Processes

This assumption of forming an abstract homogeneous continuum through the combination with probabilities is recognized by von Neumann and Morgenstern as central to their work in the following quote (von Neumann and Morgenstern 1953, p. 20):

> “We have assumed only one thing—and for this there is a good empirical evidence—namely that imagined events can be combined with probabilities. And therefore the same must be assumed for the utilities attached to them,-whatever they may be”.

In the von Neumann and Morgenstern language, “utilities” are the entities of the extensive structure. It is assumed that any entity of the expected utility theory primitive structure can be combined with probabilities. In other words, any entity of the primitive structure can be the outcome of a probabilistic process. Since probabilities must lie *beyond the control of the individual,* and that processes must remain *under the control of the individual* (see chapter 1), it results that processes cannot be combined with probabilities and are therefore excluded from the primitive structure of expected utility theory.

As an example, suppose you must attend an important conference and you realize that your passport is no more valid[44]. Unfortunately, there is not enough time left to settle the matter and you must choose between going to the conference and cancelling the trip. You expect that if the customs officer notices the expiration date, then you may have to lie to him saying that you didn’t notice. In this manner, we could say *prima facie* that lying has been combined with probabilities. However, there may remain a moral displeasure or discomfort in choosing to possibly face a situation like this. Such a displeasure or discomfort is experienced with certainty once the responsibility is taken to go to the conference. As such, it remains under the control of the individual and is a matter of process, not of consequences.

This clarifies why process preferences cannot be integrated within the structure of expected utility theory and justifies their specific consideration.

## 2.3 Embedding Expected Utility Theory in a Procedural Context

The set of axioms presented in Chapter 1 is complemented with two axioms from expected utility theory: an axiom of monotonicity and an Archimedean axiom. As such, expected utility theory remains formally identical to its original formulation. Such an embedment modifies the interpretation of expected utility theory with re-

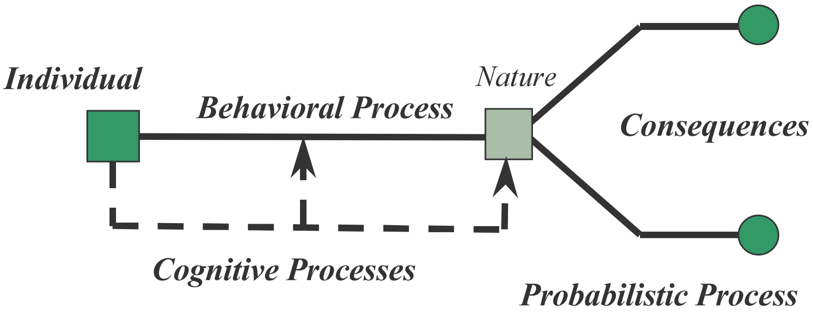
---

[44] This example was proposed to me by J. Drèze.

gards to empirical observation, since consequential preferences are no more assumed to be the empirical observable.

### 2.3.1 Axiomatic Formulation of the Model

Probabilistic consequences can be represented as in Figure 6. They correspond to the whole composed of "Nature, Probabilistic Processes, Consequences". From now on, the term consequence is meant to designate a probabilistic consequence, even if it is assigned a probability one[45].



6.Behavior as Process and Probabilistic Consequences

Recalling the model of chapter 1, a process is denoted by $a \in A$. A consequence is denoted by $c \in C$. The set $A$ is the set of *processes* and the set $C$ is the set of *consequences*. The dependence between processes and consequences is reflected by a *consequence function* $g$ with domain $A$ and range $C$. This is written $g : A \rightarrow C$, $a \mapsto g(a) \in C$. A *situation of choice* corresponds to the set of *available behaviors* $B = \{(a, g(a)) : a \in A\}$. For convenience, the four axioms presented in chapter 1 are stated again as follows.

A1. The binary relation $\succsim^B$ is a weak ordering of $B$.
A2. The binary relation $\succsim^A$ is a weak ordering of $A$.
A3. The binary relation $\succsim^C$ is a weak ordering of $C$.
A4. If $a \succsim^A a'$ and $g(a) \succsim^C g(a')$ then $(a, g(a)) \succsim^B (a', g(a'))$.

More structure is now assumed for the set $C$ of consequences, which is taken to be a *mixture set* (Herstein and Milnor 1953)[46]. This notion is stated below.

**Definition 5** *A set $C$ is said to be a* ***mixture set*** *if for any $c, d \in C$ and for any $\lambda \in [0, 1]$ we can associate another element, which we write as $\lambda c + (1 - \lambda)d$, which is again in $C$, and where*

---

[45] This does not mean that processes can be treated as consequences of probability one. The condition to be a consequence is that they "can be combined with probabilities", not that they always are.

[46] This definition replaces two axioms of von Neumann and Morgenstern about the algebraic properties of the structure (1953, axioms 3:C:a and 3:C:b) and allows probabilities to be 0 or 1.

$1c + (1-1)d = c,$
$\lambda c + (1-\lambda)d = (1-\lambda)d + \lambda c,$
$\mu[\lambda c + (1-\lambda)d] + (1-\mu)d = (\lambda\mu)c + (1-\lambda\mu)d,$
*for all* $c, d \in C$ *and all* $\lambda, \mu \in [0,1]$.

The axioms of von Neumann and Morgenstern are then[47]:

**Axiom 5** *If* $c \succ^C d$ *then* $c \succ^C \lambda c + (1-\lambda)d$, *for all* $\lambda \in [0,1]$.

**Axiom 6** *If* $c \succ^C e \succ^C d$ *then there exists a* $\lambda \in [0,1]$ *such that* $\lambda c + (1-\lambda)d \succ^C e$.

For convenience, the following definitions are recalled:

**Definition 6** *A functional* $u$ *is* ***order-preserving*** *if and only if, for all* $c, d \in C$ *we have* $c \succ^C d \Longleftrightarrow u(c) > u(d)$.

**Definition 7** *A functional* $u$ *is a* ***linear*** *function on* $C$ *if it is a real-valued function for which* $u(\lambda c + (1-\lambda)d) = \lambda u(c) + (1-\lambda)u(d)$.

**Definition 8** *Two linear functional* $u, v$ *are related by* ***a positive affine transformation*** *if there are real numbers* $\omega_1 > 0$ *and* $\omega_0$ *such that* $v(c) = \omega_1 u(c) + \omega_0$ *for all* $c \in C$.

The well-known theorem of expected utility theory is thus:

**Theorem 3** *If C is a mixture set, then axioms 3, 5, 6 hold if and only if there exists an order-preserving linear function on C. Such a functional is unique up to a positive affine transformation.*

This is rigorously standard. Proofs of the theorem can be found in von Neumann and Morgenstern (1953, A2), Herstein and Milnor (1953), Fishburn (1982), etc. This original set of axioms is never used because we usually state explicitly the independence axiom[48]. This is indeed relevant when we want to relax that axiom but is not necessary when the original structure and set of axioms remain intact.

### 2.3.2 "Real Entities" in the Theory of Expected Utility Theory

My remarks are directed towards the interpretation of the utility of an event, an object, an outcome, a consequence, etc. In their axiomatic development, von Neumann and Morgenstern made explicit that the system of "utilities" in their language was an abstract system. This is reflected in the choice of their structure and in particular in their semantic substitution of "utilities" for "real events".

Entities of the von Neumann and Morgenstern "system" are not "real entities", and no "real entities" are required to make the system formally consistent. As an

---

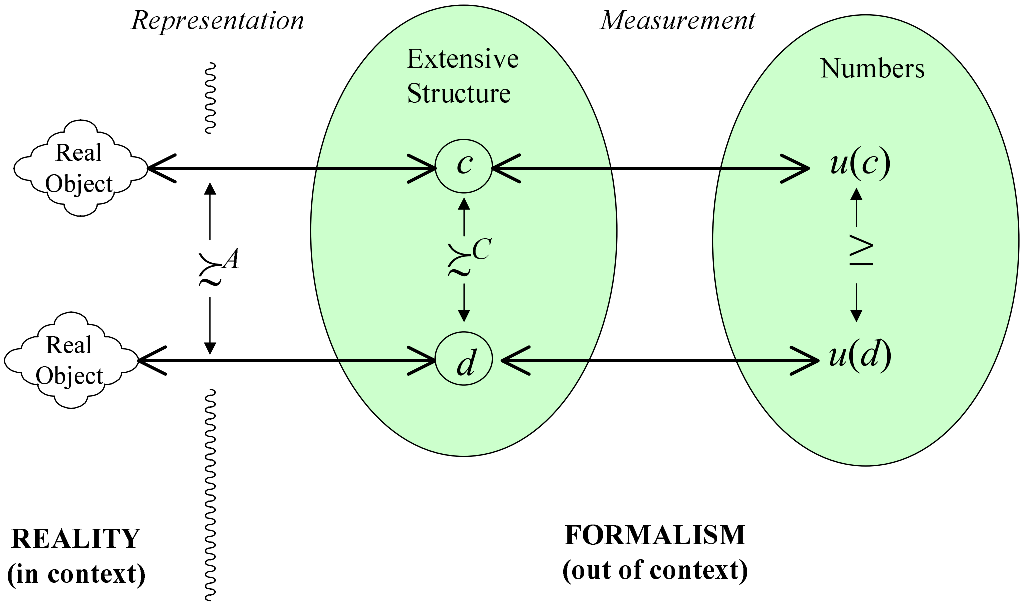
[47] In the original formulation, they are expressed for both $\succ$ and $\prec$.

[48] See Malinvaud (1952); Fishburn (1989); Fishburn and Wakker (1995) for a more detailed discussion of this point.

extensive structure, it has required the construction of a "quotient set" of equivalence classes by applying an equivalence relation to a more primitive set of real entities. This is called in mathematics an "abstraction of qualities" and corresponds to the substitution of monotonic and continuous variables for real entities. It characterizes formal models as idealizations of the reality and as such, is ubiquitous throughout all science[49]. A recent clarification due to Aumann gives other examples:

> "In any axiomatic system, the arguments depend crucially on hypothetical, artificial situations that never existed. The essence of the axiomatic approach is that it works with an entire system, and with the relations between the objects in it. In particular, it relates the given, 'real', situation to a whole lot of other situations, all of them hypothetical. This happens constantly—in Arrow's (1951) social welfare theory, in the Shapley value (1953), in Nash's (1950) bargaining solution, indeed even in Savage's (1954) development of probabilities themselves!" (Aumann 1998, p. 935).

In that respect, consequences are not real objects or events or "real entities" but "idealizations" or "concepts". We may say that they are *representations* of real entities, whatever they may be. It is relevant in the present context to make explicit such distinction between representation and measurement, which is proposed in Figure 7 (adapted from Grize 1994).



7.Representation and Measurement

A classical interpretation is for instance the following.

---

[49] See von Neumann and Morgenstern (1953, §3.4.3); Luce and Narens (1987, p. 1527); Fishburn and Wakker (1995, pp. 1132-3).

If $\mathcal{P}$ is a non-empty convex set of probability measures defined on a Boolean algebra of subsets of $X$, the set of outcomes, then it is a *mixture set*. Utility of "real outcomes", i.e. elements $x \in X$, is usually introduced through the following definition: $u(x) = u(p)$ when $p(x) = 1$ $(p \in \mathcal{P})$[50].

Such an introduction of the "utility of real outcomes" should not overlook the representation that occurs. *Ex ante*, combinations of outcomes with probabilities are not "real entities" but abstractions that exist only as states of mind of individuals. They are what the concept of an "apple" is with respect to "this now designated apple here". Indeed, any *two* apples differ in the reality, otherwise there would be only *one*. The representation abstract their singularity to consider them two "indifferent" elements of a unique class of *equivalence* called "apples". The same occurs when apples and pears are conceptualized as "fruits". The same occurs with combinations of probabilities with consequences at a more abstract level.

As such, the "utility of $x$" remains the utility of the real entity "*represented* as $x$". Whether the relation among represented entities correspond to the relation among real entities is a matter of empirical investigation. Usually, such a correspondence is assumed to hold and then tested onto the real world. All representations are a priori empirically equivalent, an assumption conveniently called "representational invariance"[51]. I am attempting to reflect this formally through the combination of the binary relation $\succsim^C$ with the binary relation $\succsim^A$.

### 2.3.3 Qualifying the Relation between the Model and the Reality it Represents

When it is said that any representation is empirically equivalent, the abstract relation among the abstract entities of the extensive structure—$\succsim^C$ in our model, is identified with an observed relation—$\succsim^B$ in our model. The novelty here is to make such distinction explicit and meaningful, and this is possible because the structure allows these two types of preferences to be formally distinct. More precisely, the way $\succsim^C$ "expresses" empirically now explicitly depends on $\succsim^A$ . Therefore, a correspondence between the abstract relation $\succsim^C$and the observable $\succsim^B$ necessitates the explicit assumption of *neutral representation*: $a \sim^A a'$ for all $a, a' \in A$. In conventional formal models I know of, this hypothesis seems to be always assumed beforehand in an implicit manner.

From these, our model takes two directions. First, it is assumed that the neutral representation assumption is legitimate in particular experimental settings. Second,

---

[50] In Savage (1954), the primitive entities are functions from states to consequences called acts. A similar procedure is then used to define utility of consequences by postulating the existence of "constant acts".

[51] See for instance Fishburn (1988, p. 27) where this is identified as a general principle of both expected and non-expected utility models. The issue may be that individuals have judgments on such representations, as "frames" or "ways" to look at the world. See Tversky and Kahneman for a systematic study of the influence of framing effects. A nice conceptual investigation about the rationality of frames is Fillieule (1996).

it is proposed to use this knowledge of the abstract relation $\succsim^C$ in order to qualify how the empirical relation $\succsim^B$ deviates from it.

In that respect, the definition of $u(x)$ given page 39 is now *explicitly* an *abstract* definition of "the utility of the outcome $x$ independently of the process to reach it". It is indeed a *judgment*, not an *observation* since it is revealed *dependently* of a process[52]. Without $\succsim^A$, there is no "hurdle" to prevent the intuition to assign the quantity measured to the real object. Formally speaking, it nevertheless remains solely assigned to a particular choice of *attributes* to describe the real object. Such choice of *attributes* characterizes the idealization of the real entity, i.e. its representation as a member of an equivalence class. The empirical correspondence can be established only when such representation is neutral.

Supposing that any entity—on which preferences apply—is "fully" described may not solve the above issue without raising other difficult ones. By "fully" describe, I refer to the idea that one may increase the number of *attributes* so that no *attribute* is left out from the measurement. The logical difficulty is this: to be measured, any attribute must be possessed by all entities. However, that two entities have *all* their attributes in common requires again an abstraction. Indeed, if no attribute distinguishes two real entities, then there are no two real entities but only one. And if there are two distinct entities, then they must be distinguished by at least one attribute: themselves. This argument corresponds to the distinction between extensive and intensive—or intentional—definitions. Extensive entities are always composed of other entities. On the other hand, entities defined intentionally cannot be composed and remain singular.

As an example, the mass of an object can be measured by a number. Such a number does not measure the object itself, but the whole equivalence class of objects deemed to have the same mass. The number is thus assigned to the mass of the object, as an equivalence class comprising the object itself but not only *this one*. Whether such measure can be empirically validated with a particular object depends on the possibility of "reproducible observations" (see the quote given page 30). This in turn corresponds to a form of invariance of the measuring device whose behavior must reflect properties of objects always in the same manner, i.e. independently of the context. This point is that individuals are the measuring devices themselves in the case of rational behavior. Instead of assuming that individuals are always the same, we may be interested in how they change and evolve. This is what the model does through the relation $\succsim^A$ . Instead of being "hidden", the preference relation $\succsim^A$may reflect the quality of the description or the quality of the experimental investigation like a sort of "qualitative" type of "error".

## 2.4 The (dis)Utility of Gambling and the Allais Paradox

The following attempts to develop these exploratory considerations in a more concrete manner. After a brief review of the literature on the (dis)utility of gambling, it

---

[52] And it can be measured on an interval scale depending on the assumption of neutral representation.

is shown how such an attribute cannot be possessed by all entities of the extensive structure and thus, cannot be integrated *within* the von Neumann and Morgenstern framework. It is then proposed to leave it as one argument, *outside* the utility function, and to infer from choice a ranking over such an argument. Some examples of predictions involving a process utility are also proposed.

### 2.4.1 A Brief Review of the Literature on the Utility of Gambling

It has long been recognized that people can find pleasure in the mere act of gambling (Pascal, 1670). It is only when von Neumann and Morgenstern proposed their axiomatic approach to expected utility theory that *utility of gambling* became a puzzle for researchers interested in the formalization of rational choice. Von Neumann and Morgenstern show that their axiomatic treatment "eliminates the specific utility or disutility of gambling" (1953, p. 629, see also pp. 28, 632).

Marschak (1950, p. 138) takes the example of a climber who would prefer a lower probability of severe injury, but would at the same time not go climbing if such risk was null. Typically, this climber would have a utility of gambling—or a taste for riskiness—that cannot be taken into account by his utility function.

It progressively became conventional to exclude any utility or disutility of gambling per se from expected utility theory, and to consider such motives "irrational" (e.g. Marschak 1950, p. 139, Harsanyi 1978, p. 225). Luce and Raiffa (1957, p. 26) associate the exclusion of the utility of gambling to the reduction of compound lotteries assumption. Fishburn (1980) explores the issue by proposing a model that distinguishes "sure outcomes" from "non-degenerate and not-compounded lotteries". This allows to express a utility of gambling as an additive attribute although the model remains limited in its descriptive capabilities. Pope (e.g. 1995 1998) dedicates several articles to the utility of gambling, enlightening its historical, theoretical and practical relevance. She questions that some motivations or satisfactions should be excluded from the formalization of rational choice because they are too "emotional" and thus impede an "objective" approach to rationality (1998, p. 103). She also proposes the idea of considering the utility of gambling as resulting from the presence of a "pre-outcome period" (1995)[53]. The problem remains to formalize such type of intuitive argument and to broaden the von Neumann and Morgenstern framework accordingly. No satisfying solution seems yet to have been found to include the utility of gambling into the expected utility framework. Fishburn (1989, p. 153) expresses this opinion, as well as Morgenstern himself:

> "I want to make it absolutely clear that I believe—as von Neumann did—that there may be a pleasure of gambling, of taking chances, a love of assuming risks etc. But what we did say and what I do feel I have to repeat even today after so many efforts have been made by so many learned men, is that

---

[53] This is somehow close to the notion of process developed here. Pope however distinguishes two distinct stages. I have argued instead that reaching a consequence already implies a process. In that manner, even one stage is dual. It is composed of a *relation* (1) to a *substantive entity* (2). This is why the model is structured as it is.

> the matter is very elusive. I know of no axiomatic system worth its name that specifically incorporates a specific pleasure or utility of gambling with a general theory of utility" (1979, p. 181).

The result of this inadequate treatment is that a specific utility or disutility of gambling violates the expected utility normative framework, it creates a "bias" (e.g. Allais 1953, Kahneman and Tversky 1979). Attempting to remedy these descriptive inadequacies, I have evoked how non-expected utility approaches propose a non-linear monotonic functional form that remains consistent with "irrational" behaviors, like for instance in the Allais paradox (see for instance: Machina 1982). However, these approaches rely on transformations of probabilities and do not specifically incorporate a specific utility of gambling. Moreover, the independence condition between probabilities and consequences is relaxed and the formalism proposed thereby loses the mathematical tractability of the von Neumann and Morgenstern approach.

Our approach differs from these attempts at modeling a utility of gambling by explicitly leaving the von Neumann and Morgenstern framework intact, so as to preserve its elegance and intuitive appeal. Gambling is indeed introduced as a process attribute, not as an attribute of consequences. Such procedural preferences embed expected utility theory in a larger framework.

### 2.4.2 Demonstrating the Consequential Nature of Expected Utility Theory

It is sometimes argued that constructing more "elaborated" consequences is always possible (e.g. Samuelson 1952, Hammond 1988, 1996 denies that a consequentialist reduction leaves any "gap"). Pope has long and persuasively argued that this is not correct. I show here how treating *Gambling* and *Not Gambling* as attributes of consequences leads to *logical difficulties* that prevent rigorous reasoning. It illustrates the argument that consequences cannot encompass "everything".

The consequential nature of expected utility theory stems from the reduction of preferences to preferences for outcomes or, in the words of Fishburn (1988, p. 27), from a "*reduction principle*" stating that:

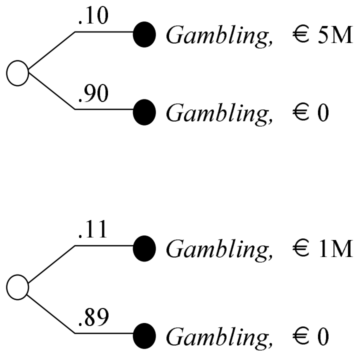
> "for comparative purposes of preference and choice in risky decisions, it suffices to characterize each alternative in terms of its probability distribution over potential outcomes".

This reduction is indeed implicit for both expected and non-expected utility approaches. It dates back to Bernoulli (see 1954). Recall the quote from von Neumann and Morgenstern stating that it must be possible to combine utilities with probabilities (see the quote page 35). Utility of gambling illustrates how such axiomatic approach restricts itself to primitive entities that *can* be combined with probabilities. The arguments runs as follows:

> The act of *Not Gambling* cannot be combined with probabilities, since com-

bining *Not Gambling* with a probability becomes *Gambling*.

This demonstrates that the attribute *Not Gambling* is excluded from von Neumann and Morgenstern *"utilities"* because of the mere choice of representing primitive entities as combinations of probabilities with utilities. The argument for the attribute *Gambling* can be carried out using a multi-attribute version of expected utility theory, where consequences are composed of several attributes (Keeney and Raiffa 1976, see also Fishburn 1982). The following reasoning shows that treating *Gambling*—denoted $G$—as one attribute of consequences is not logically "meaningful". Consider for instance the choice presented in Figure 8, which is one of the choice situation in the paradox of Allais (1953).



8.Gambling as an Attribute of Consequences

Treating $Gambling$ as an attribute ($G$) of consequences would lead to express a preference for the upper probabilistic consequence as

$$.10u(G, 5\mathrm{M}) + .90u(G, 0) > .11u(G, 1M) + .89u(G, 0) \tag{2.5}$$

where $u$ is a monotonic linear multi-attribute functional. Hence, we can re-write the above inequality as

$$.10u(G, 5M) + .89u(G, 0) + .01u(G, 0) > .11u(G, 1M) + .89u(G, 0). \tag{2.6}$$

A substitution of a common consequence leads equivalently to

$$.10u(G, 5M) + .89u(G, 1M) + .01u(G, 0) > .11u(G, 1M) + .89u(G, 1M) \tag{2.7}$$

$$\Leftrightarrow .10u(G, 5M) + .89u(G, 1M) + .01u(G, 0) > 1u(G, 1M). \tag{2.8}$$

The entity $1u(G, 1\mathrm{M})$ has however no possible meaning.

It states that the individual receives with probability 1 an amount of €1 Million with the attribute $Gambling$. But the meaning of $Gambling$ is to characterize a $probabilistic$ consequence, not a consequence attained with probability one[54].

Now it is clearer that such attributes cannot be included within expected utility theory, the following section introduces the model with processes.

### 2.4.3 Modeling a Process (dis)Utility of Gambling

The interpretive context is set-up as follows: processes take the attribute $Not\ Gambling$ ($\overline{G}$) if it leads to a consequence with probability one and it takes the attribute *Gambling* ($G$) otherwise. An individual who has a preference for $G$ over $\overline{G}$, i.e. for whom $G \succ^A \overline{G}$, is said to have a *utility of gambling*. An individual for whom $\overline{G} \succ^A G$ is said to have a *disutility of gambling*. This is made formally explicit by stating that $A = \{G, \overline{G}\}$. A behavior is thus written $(G, c)$ or $(\overline{G}, c)$ depending on its consequence (probabilistic or deterministic). As such, this specific application consists in leaving $Gambling$ or $Not\ Gambling$ as one argument *outside the utility function*, i.e. without an additive model on the attributes of the utility function (see Fishburn 1980 for such a model).

As such, the singular processes that lead to consequences are replaced by "labels" $(G, \overline{G})$ whose assignment is—in this case—objectively given by the situation of choice. Such a choice of process attributes consists in abstracting singularity of the processes such as to consider only their attribute $Gambling$ or $Not\ Gambling$ as influencing rational behavior. It explicitly assumes that processes possessing the same attribute are equivalent and thereby restricts the *interpretive context* of process preferences. The justification as well as the *intensity* of such process utility may or may not depend on the consequences under consideration and these features are not modelled here. It may indeed vary from one individual to the other and across contexts. This model is precisely a tool to reveal such individual specificities rigorously. Within a particular context, it is nevertheless assumed that a particular individual does not switch from a utility of gambling to a disutility of gambling. Otherwise, it is not possible to make any prediction.

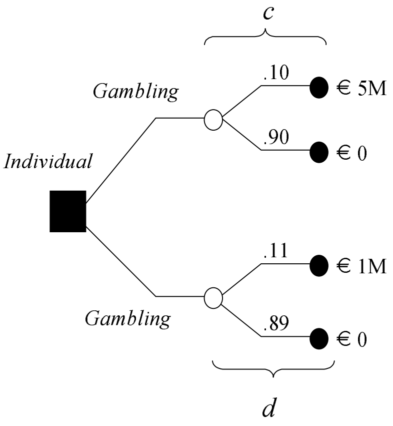
Naturally, observed behavior does not depend on such a choice. It may be that such a choice of attributes does not describe correctly the behavior of a particular individual. For instance, the model may reveal that in one choice, the individual *must* have a *utility* of gambling, while in another he *must* have a *disutility* of gambling. In such a case, we can conclude towards irrationality or, since the model always remains *open*, we can try to find another set of process attributes that preserve the assumption of rationality. In this manner, the model proposes a rigorous framework to explain behavior descriptively.

---

[54] A further inquiry into this logical difficulty may indeed be enlightening about the logical construction of expected utility theory.

### 2.4.4 Revealing a Process Disutility of Gambling in the Allais Paradox

Consider the choice situation depicted in Figure 9.



9.Gambling as an Attribute of the Process

Compare it first to Figure 8 where only consequences are depicted. Now, the *individual* is also depicted, providing intuition that the process is a *relation* between the individual and the consequence. As already said, it is distinct from the consequence although both depend on each other for their conjoint realization. Moreover, *Gambling* now characterizes the *process.* Denote $c$ the probabilistic consequence of receiving €5M with probability .1 and nothing with probability .9. Denote $d$ the probabilistic consequence of receiving €1M with probability .11 and nothing with probability .89. The empirical observation of a choice for the upper behavior can be written

$$(G, c) \succ^B (G, d). \tag{2.9}$$

Which, by the dependent revelation of consequential preferences

$$c \succ^C d. \tag{2.10}$$

In this manner, the process attribute *Gambling* has been "abstracted" and the conventional expected utility reasoning can apply. We have equivalently

$$u(c) > u(d) \tag{2.11}$$

or

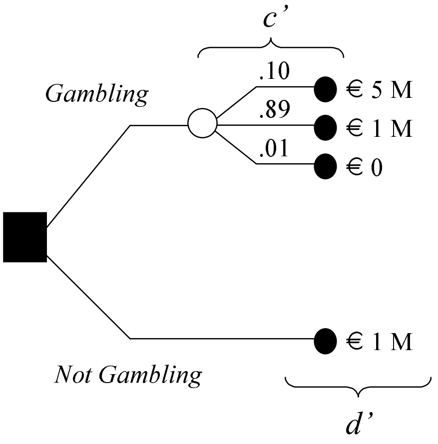
$$\Longleftrightarrow .10u(5\mathrm{M}) + .90u(0) > .11u(1\mathrm{M}) + .89u(0). \tag{2.12}$$

and therefore

$$\Leftrightarrow .10u(5\text{M}) + .89u(1\text{M}) + .01u(0) > u(1\text{M}). \tag{2.13}$$

The probabilistic consequence of receiving €5M with probability .1, €1M with probability .89 and nothing with probability .01 is further denoted $c'$, while the deterministic consequence of receiving €1M is denoted $d'$. Expression (2.13) states

$$c' \succ^C d'. \tag{2.14}$$



10.Revealing a Process Disutility of Gambling

In Figure 10, the *consequence* $c'$ of the upper behavior must be consequentially preferred to the consequence $d'$ of the lower behavior. It does *not* follow from expression (2.14) that a rational individual who has chosen the upper behavior in Figure 9 *should* choose the upper behavior in Figure 10. If the individual has a process *disutility of gambling*, choosing the lower behavior of Figure 10 may still be rational because the process of the lower behavior has the attribute *Not Gambling* while the process of the upper behavior has the attribute *Gambling*. Indeed, empirical observation reveals such process utility.

Consider an individual who has chosen the upper behavior of Figure 9 and who chooses the lower behavior of Figure 10. This is written

$$(\overline{G}, d') \succ^B (G, c'). \tag{2.15}$$

Combined with expression 2.14, this *reveals* that the individual has a *disutility* of gambling.

The argument goes as follows: if the individual had a *utility of gambling*, then, by the optimality axiom, we would have observed a choice for the upper behavior $(G, c')$. Since we observe $(\overline{G}, d') \succ^B (G, c')$, then the individual must have a *disu-*

*tility of gambling*. Note that the derivation of $c' \succ^C d'$ is possible because expected utility theory is assumed to be a valid measure of the utility of *consequences*. Otherwise, the reasoning applied here would not be possible.

For clarification, it should be observed that the observation of $(G, c') \succ^B (\overline{G}, d')$ when $(G, c) \succ^B (G, d)$ has been observed in the situation of Figure 10 is consistent with a utility of gambling, but also with a disutility of gambling. Since the degree or intensity of process preferences is not part of the model, the individual may have a disutility for gambling that has not enough "strength" to reverse expected utility of consequences. In this pattern, utility or disutility of gambling remains unrevealed without additional empirical observations.

Recall Marschak's climber example. There is nothing irrational about an individual who likes to climb, but does so in a prudent manner. For instance, a climber may be "risk-averse" while still not go climbing when such risk is null. The intrinsic preference for the presence of risk is treated as a process attribute. It reflects a process utility of gambling, as a taste for riskiness. In this manner, climbers can be said to be rational in their behavior if one considers that they are motivated by the process of climbing, while still caring about dangerous consequences. It may be that the same individual shows a disutility of gambling in a laboratory experiment while he shows a utility of gambling when climbing. The two contexts are indeed very different.

In either cases—disutility or utility of gambling—, the monotonicity of the functional form measuring expected utility of consequences *is not violated*, contrarily to the interpretation of Marschak[55]. It could even be said that an individual who prefers a probabilistic consequence to a deterministic consequence that is better than any of the outcomes of the probabilistic consequence is rational. Such a behavior would reveal a "particularly strong" utility of gambling—as a taste for *wonder*—similar to the preference an individual may have for the suspense of opening a gift.

With this model, patterns of behavior that are systematically observed do not appear "irrational" but are rather rationally explained *ex post* by a process utility. The conditions under which a process utility can be a rigorous explanation are formally constrained by the model. The interpretation of such process utility, for instance as a process utility of gambling, is made explicit through the specification of the process attributes.

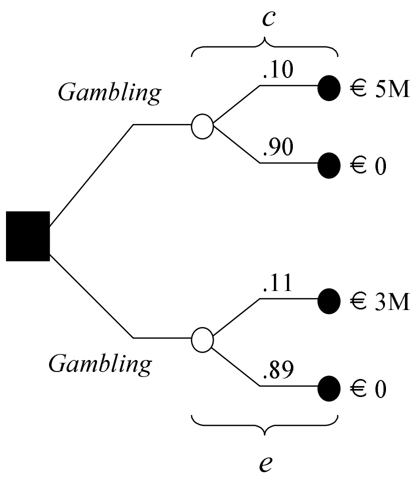
The following section shows that the proposed approach also allows predictions *ex ante*.

### 2.4.5 Predictions involving a Process Utility Hypothesis

Suppose that we have observed an individual choosing the upper behavior in Figure 9, and then the lower behavior in Figure 10. As argued, it reveals a process disutility of gambling. Within the same context of choice, consider now the situation depicted in Figure 11.

---

[55] The utility increases when the probability of severe injury decreases but decreases when such a probability becomes zero.



11.Dominated Consequences

Compared with Figure 9, the consequence of the lower behavior, denoted $e,$ has been improved such as to include a .11 probability of receiving €3 Million. Suppose that we observe a choice for the lower behavior:

$$(G, e) \succ^B (G, c). \tag{2.16}$$

The dependent revelation of consequential preferences leads to

$$.11u(3\text{M}) + .89u(0). > .10u(5\text{M}) + .90u(0) \tag{2.17}$$

or equivalently

$$u(3\text{M}) > .10u(5\text{M}) + .89u(3\text{M}) + .01u(0). \tag{2.18}$$

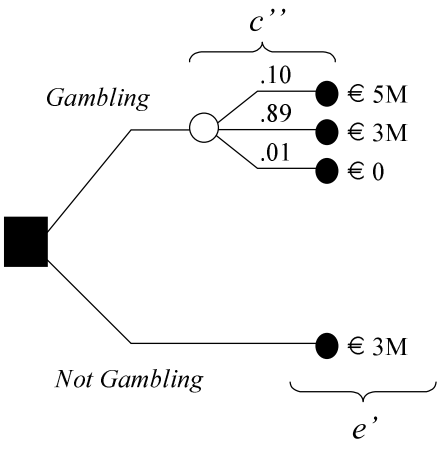
Denoting $e'$ the deterministic consequence of receiving €3M and $c''$the probabilistic consequence of receiving €5M with probability .1, €3M with probability .89 and nothing with probability .01, we have

$$e' \succ^C c''. \tag{2.19}$$

Given the individual's disutility of gambling, we can use *optimality* to predict in the situation of Figure 12 that a rational individual should choose the lower behavior:

$$\{\overline{G} \succsim^A G \text{ and } e' \succ^C c''\} \Longrightarrow (\overline{G}, e') \succsim^B (G, c''). \tag{2.20}$$

If it is observed instead that $(G, c'') \succ^B (\overline{G}, e')$, then the explanation of behavior with a utility or a disutility of gambling is contradicted and the individual must be considered "irrational".



12.Predicting Behavior with Process Preferences

Therefore, a falsifiable prediction for the situation of Figure 12 can be proposed based on a pattern of behavior for the situations of Figure 9, 10 and 11. This prediction cannot be made without considering the specific influence of process utility since behavior in the situations of Figure 9 and 10 falsify expected utility theory if not properly restricted to consequences. Treating choices of Figure 9 and 10 *and* choices of Figure 11 and 12, this model can interpret a larger class of behaviors[56]. As for its descriptive ability, the normative ability of the model builds on expected utility theory rather than invalidates it.

For clarification, suppose finally that we observe a choice for the upper behavior in Figure 11, i.e. we observe $(G, c) \succ^B (G, e)$. This implies that $c \succ^C e$. Through expected utility theory, this is equivalent to $c'' \succ^C e'$. Since previous observations have revealed a disutility of gambling, optimality cannot be invoked to predict the behavior of the individual in the choice situation of Figure 12. Again, rational behavior depends on the strength of the process disutility of gambling.

By explaining systematic violation of expected utility of consequences in the presence of certainty through process utility, it is thus possible to build on expected utility theory to improve our understanding of rational behavior. Expected utility theory helps to reveal what it misses because it formally “draws the line” between consequential and procedural considerations. The nature of the process dimension however prevents such a revelation to be of quantitative nature. The “degree or “intensity” of process utility is not modelled and it remains *ordinal.* As a result, the predictive ability of the model, like its descriptive ability, is limited accordingly.

Although such an ordinal measurement of process preferences may be improved, it is certainly not a “weakness” of the model in comparison to expected utility theory. Conventional interpretations of expected utility theory rely on the implicit as-

---

[56] Besides neutrality and optimality, a consequence of a preferred behavior can always be replaced by a better one (this is a form of dominance).

sumption that individuals do not have preferences for processes per se (neutrality of representation). With that assumption, the proposed model is equivalent to expected utility theory and is thus not weaker than it.

## 2.5 Additional Remarks on the "Peculiarities" of the Model

The consideration of a utility of gambling thus may induce a bias in the measurement of utility for consequences. Nevertheless, utility of these consequences can still be measured through comparisons of non-degenerate probability distributions[57].

The attainment of a quantitative measurement of consequential preferences, without a similar result for processes, leaves us with a peculiar "numerical structure" composed of a quantitative scale (an interval one) and a qualitative scale (an ordinal one) that combine through the optimality axiom.

As such, choice orders dual entities as, for example

$$\begin{gathered}\text{“5 units of the first rank”}\\ \succsim^{B}\\ \text{“3 units of the first rank”}\\ \succsim^{B}\\ \text{“3 units of the second rank”}\\ \succsim^{B}\\ \text{“1 unit of the second rank”.}\end{gathered}$$

The "*open*" characteristic of the formalism explored in this dissertation thus becomes clearer when observing choice over dual entities as

$$\begin{gathered}\text{“3 units of the first rank”}\\ \succsim^{B}\\ \text{“5 units of the second rank”.}\end{gathered}$$

Although choice empirically orders such entities, an abstract ordering of them is missing. This characteristic of the model can be related to the remark of chapter 1 (section 1.7 page 22)about the formal distinction between *ex ante* and *ex post* perspectives[58]. Later, section 3.4.6 proposes to reflect such a structure by multiplying an ordinal utility function on processes with a cardinal utility function on consequences.

Another critical issue is whether and how a more elaborated measure of such type of utility can be obtained. Although a "population" approach may enhance such measurement (the same individual across time or different individuals at the

---

[57] See for instance Wakker and Deneffe (1996) and their references. See also Fishburn (1970) and Krantz et al. (1971).

[58] Further investigations are required in this respect. This combination of ordinality and cardinality has been influenced by the ideas of Piaget and Grize about the construction of natural numbers (see for instance 1967 and in particular the work of Grize on the construction of natural numbers pp. 512-525).

same time). A key point is whether the influence of emotional tastes can be assumed to be qualitatively similar across the population (Kahneman et al. 1997, p. 380).

## 2.6 Conclusion to Chapter 2

Expected utility theory has been widely criticized for its descriptive inadequacy. Notwithstanding the massive efforts to construct alternative approaches, the dominant view is that it remains the proper normative model when facing probabilistic consequences.

Rather than modifying its axioms, this approach proposes to *embed* the von Neumann and Morgenstern approach in a procedural context.

The main arguments of the chapter may be summarized as follows:

- Consequential preferences can be quantitatively measured to the extent these entities can be *combined* with *probabilities*. This is standard expected utility theory.
- If processes necessarily remain *under the control of the individual*, then they cannot be combined with probabilities. As a result, preferences for processes are excluded from expected utility theory.
- Preferences measured through expected utility theory are *judgmental*. They do not necessarily reflect observation, except when representation is assumed to be neutral. Such an assumption is a matter of appropriate experimental setting.
- Expected utility theory is not violated if properly restricted to consequences and the formalism of expected utility theory remains intact.
- Expected utility theory is necessarily consequential and provides *a formal separation of processes and consequences*.
- Unless a type of *operation of combination* can be assumed among *processes*, they remain *singular*. As a result, process preferences cannot be quantitatively measured. They are revealed *ordinally* as a *ranking* over processes.
- Through an explicit choice of *attributes*, the interpretation of revealed process preferences can be made *rigorous* and subject to *contradiction*.

In the next chapter, the relation between processes and consequences is again different. Consequences result from the combination of processes taken by different individuals. This is the framework of game theory.

# Chapter 3
# Processes and Interdependent Consequences

> *"It is that the mathematical properties of a game, like the aesthetic properties, the historical properties, the legal and moral properties, the cultural properties, and all the other suggestive and connotative details, can serve to focus the expectations of certain participants on certain solutions".*
> *Thomas C. Schelling*

## 3.1 Normative and Descriptive Interpretations in Game Theory

Game theory faces interpretive difficulties (Rubinstein 1991). Since this theory has many formulations and applications, these difficulties range over various issues. This chapter focuses on the distinction between normative and descriptive interpretations of the simplest form of strategic games. Normatively, solution concepts apply to games where only consequences are considered. Descriptively, rational behavior is nevertheless influenced by process preferences. Instead of treating any behavior that deviates from the normative solution as "irrational", we show that the notion of process preferences helps to interpret departures from consequential rationality. In particular, it sheds light on how a social norm, as a profile of process preferences, helps individuals to coordinate on the best behavior for all.

When a behavior is observed, it is *descriptively* interpreted as rational and preferences are revealed. Then, these preferences are *normatively* assumed to follow some axioms and solution concepts. This allows other hypothetical rational behaviors to be deduced. Then, if behavior that is not consistent with the deduced rational behaviors is observed, it must be treated as *"irrational"*. Descriptively, the individual didn't act according to his preferences or alternatively, his preferences are not consistent with the normative model. This pattern constitutes a "paradox". It relies on two distinct interpretations of rationality: a descriptive one (individuals do what they prefer) and a normative one (individuals behave consistently).

As a typical example, cooperative behavior in the Prisoners' Dilemma is consequentially irrational: normatively, a rational individual should defect. On the other hand, cooperative behavior in the Prisoners' Dilemma is empirically observed: descriptively, a rational individual may cooperate[59].

---

[59] Sally (1995) provides a meta-analysis on Prisoners' Dilemma experiments. One may like to consult Rapoport and Chammah (1965); Rapoport et al. (1976); Rapoport (1988). Note that most recent empirical studies have focused on repeated games.

The difficulty is one of conceptual nature and of precision more than of technical sophistication. The reduction of preferences to consequential preferences is an implicit assumption that is quite systematic in game theory. Besides, notational and analytical practices tend to preclude the distinction between the game of consequences and the real game acted by the individual. For these reasons, the following argument is presented in its simplest form. It focuses on the simplest version of the game: one-shot (individuals act only once), simultaneous moves (they don't know what the other does), no randomization (they really have to choose one behavior), perfect information (there is no "uncertainty" considered). This expression of the argument in this minimal setting should render more salient how process preferences provide an intuitive explanation to the Prisoners' Dilemma paradox. Although the present state of the theory considers that such a game has only one normatively rational solution: *defection*, our model has different rational solutions that depend on individuals themselves and on the consequences at hand.

## 3.2 A Brief Review of the Literature

As a one-shot game approach, our model must be distinguished from the *dynamic* developments of game theory, in particular concerning the Prisoners' Dilemma. It is indeed possible to construct an argument according to which cooperation appears rational in a repeated setting of this game. This nevertheless implies a modification of the basic rationality assumption that individuals "act for their own good". Individual themselves may even "disappear" when models consider only populations[60]. Although these dynamic approaches are instructive, they cannot solve the initial paradox because they rely on repetition. Cooperation is indeed found in one-shot games. Otherwise, it would not be a "paradox" and would not have attracted so much attention.

Because it maintains the assumption "one individual, one choice", the approach relying on *fairness* is closer to our subject. If cooperation is *fair*, and individuals *prefer* to be fair, then indeed cooperation may be preferred *ex ante*. *Ex post*, it may also pay. In such models, individuals do not act because of the consequences of their actions, but as guided by deontological considerations. As an example, Gauthier (1986) "solves" the Prisoners' Dilemma in this manner and shows that cooperation is "the rational solution". As for dynamic settings, individuals do not really "act for their own good" in that approach either. They choose some "dispositions" and the game of dispositions is not a Prisoners' Dilemma anymore. Moreover, it is difficult to accept that it would *always* be rational to cooperate. A Prisoners' Dilemma where you loose €1 M if you cooperate when the other defect may not be rationally played in the same manner than one where you loose only €1, everything else being equal. The approach of Gauthier may be a moral ideal, it does not however make the link with descriptive investigations. Although cooperation is *sometimes* observed, it is

---

[60] I do not discuss in details this point, Aumann (1997) does it with clarity. Axelrod (1984) is one example of such approaches. See also Kandori (1992); Fudenberg and Levine (1998).

certainly not *the* rational solution, since it surely depends on the individuals and on the consequences.

A recent attempt due to Rabin (1993) links consequential considerations with ethical and emotional concerns[61]. Rabin considers that individuals are willing to sacrifice their own material well-being to help those who are being kind and punish those who are being unkind. In this manner, Rabin compares "*psychological games*" from games of consequences, that he calls "*material games*". He then studies how equilibria of material games can be related to equilibria of psychological games. This approach however does not directly combine outcomes and emotional concerns. In Rabin's own words:

> "Could these emotions be directly modeled by transforming the payoffs, so that one could analyze this transformed game in the conventional way? This turns out to be impossible" (ibid., p. 1285).

There are indeed difficulties related to the question of how games can be transformed by the context in which they are played. It amounts to include in consequences properties that are excluded by construction. This may be why he thinks it is "impossible"[62].

The failure of one particular means need not exclude the possibility of achieving the same end by another means. My contention is that the theoretical framework of game theory already contains a "natural" entity which allows for ethical or emotional concerns to be formally expressed. The entity that I want to consider for the support of ethical or emotional concerns is the *process* of behavior itself, called *actions* in game theory. As developed along this dissertation, the mere modification that is proposed is to consider that individuals have preferences over actions per se, as processes to reach consequences. So we don't need to add any new entity to the consequential framework of game theory. We merely have to consider that preferences may not be reduced to preferences for consequences. In this manner, cooperation in a Prisoners' Dilemma can be consistent with rationality as the choice of the preferred behavior. Moreover, process preferences shed light on the process by which individuals reach an equilibrium. A simple definition of social norms as profiles of process preferences can be proposed and a refinement of the Nash Equilibrium concept—called an Optimal Nash Equilibrium—characterizes situations where a social norm influences behavior towards the best consequences for all. An Optimal Nash Equilibrium, when it exists, is always Pareto-efficient. It may be interpreted as a "focal equilibrium" (Schelling 1960) and the social norm that ensures such equilibrium may correspond to a "convention" (Lewis 1969).

The rest of the chapter is organized as follows: section 3 recalls the formulation of strategic consequential games in an attempt to avoid semantic and notational misunderstandings. Section 4 presents the definition of rationality in the context of

---

[61] See also Binmore (1994) who nevertheless tend to reduce everything to the same type of considerations.

[62] On psychological games, see Gilboa and Schmeidler (1988) and Geanakoplos and Pearce (1989).

game-theoretic interactions and proposes appropriate solution concepts. Section 5 applies our model to the analysis of the Prisoners' Dilemma while section 6 concludes.

## 3.3 Consequential Games

This section is merely a reminder of basic notions of game theory. It is adapted from Osborne and Rubinstein (1994).

### 3.3.1 Set-up

Consider a set $N$ of $n$ *individuals*. For each individual $i \in N$, consider a set $A_i$ of *processes* or *actions*: $a_i \in A_i$. The set of *profiles of processes* is denoted $A = \times_{j\in N} A_j$. A profile of processes thus consists in one process per individual. We note $a = (a_i, a_{-i}) \in A$, where $-i = \{j \in N \ : j \neq i\}$. Furthermore, consider a *consequence function* $g$ of domain $A$ and of range a set of profile of *consequences* $C$. We note $g : a \longmapsto g(a) \in C$. Such a consequence function assigns a profile of consequences to any profile of processes, i.e. the same profile of processes can be assigned different consequences depending on individuals. In terms of notation, we write $C = \times_{j\in N} C_j$. If we desire to make explicit the consequence for the individual $i$ of the profile of processes $a = (a_i, a_{-i})$, we can write $g_i(a) = c_i \in C_i$ which denotes the $i^{th}$ component of the profile of consequences $g(a)$. We consider further that each individual $i$ has a *preference relation* $\succsim_i^C$ over consequences[63].

### 3.3.2 Definition of a Consequential Game

A consequential game can be summarized by the following definition.

**Definition 9** *A* ***consequential game*** $\Gamma = \langle N, (A_i), g, (\succsim_i^C)\rangle$ *is defined by*:

*(i) A finite set $N$ of individuals.*

*(ii) Non-empty sets of processes $A_i$ for each individual $i \in N$.*

*(iii) A consequence function $g$ from $A = \times_{j\in N} A_j$ to $C$, the set of profiles of consequences.*

*(iv) A preference relation $\succsim_i^C$ over $C_i$ for each individual $i \in N$.*

Processes and profiles of processes are thus not the same. When an individual implements a process, he does not control the consequence he will reach. This depends indeed on what other individuals do and renders game theory interesting. Therefore, there is already a distinction between what the individual controls (his action or process) and what the individual expects (ex ante) and obtains (ex post). It is crucial to make more explicit these two dimensions and to avoid dropping the consequence function when expressing preferences over consequences, i.e. avoiding

---

[63] I do not write $\succsim_i^{C_i}$ which would be redundant.

expressions like $(a_i^*, a_{-i}) \succsim_i^C (a_i, a_{-i})$ where the $g$ is omitted, when we do not want to assume that individuals are neutral with regards to processes[64].

I now turn on the standard paradox of the Prisoners' Dilemma.

### 3.3.3 The Paradox of the Prisoners' Dilemma

In the Prisoners' Dilemma game, two individuals can implement two processes, denoted for instance $Co$ and $De$. A pair of processes—one for each individual—generates a pair of consequences, one for each individual. We write $g_i(De, Co)$, $g_i(Co, Co)$, $g_i(De, De)$, and $g_i(Co, De)$. The process taken by $i$ is written first. Then a Prisoners' Dilemma game is characterized by $g_i(De, Co) \succ_i^C g_i(Co, Co) \succ_i^C g_i(De, De) \succ_i^C g_i(Co, De)$ for each individual $i$. Therefore, each individual is better off playing $De$ when the other plays $Co$; and each individual is better off playing $De$ when the other plays $De$. We don't need the solution concept of a Nash Equilibrium in a consequential Prisoners' Dilemma: an individual is always better off by implementing the process $De$.

If an experiment is carried out according to the above and an individual implementing $Co$ is observed, his behavior must be considered "*irrational*". What is meant by irrational is that the individual doesn't act according to his preferences. Because if he acts according to his preferences, then we must either have $g_i(Co, Co) \succ_i^C g_i(De, Co)$ or $g_i(Co, De) \succ_i^C g_i(De, De)$, which is then contradicting $g_i(De, Co) \succ_i^C g_i(Co, Co) \succ_i^C g_i(De, De) \succ_i^C g_i(Co, De)$ .

Two last remarks are in order. First, preferences over consequences can be revealed descriptively before the game is played. When we need only ordinal preferences, it suffices to take increasing amounts of money and to suppose that individuals prefer more money than less. Second, nothing in the formalism specifies the experimental context. Therefore, there should be no difference whether individuals can communicate or not before playing, know each other, belong to the same social community, etc. Of course, this makes empirical differences but the point is that such dependence is not part of the simplest version of the theory. A motivation for the following model is precisely to integrate these considerations in the basic model.

## 3.4 Social Games

Process preferences may modify the consequential game from the point of view of the individual who acts. Since processes are defined when the game is constructed, they could not have been measured beforehand, i.e. before the construction of the game. This differs from consequential preferences which do not need the game itself to be defined. The extent to which a game of consequences is modified or transformed depends the social context. I propose to refer to *social games* to designate the game in its context, i.e. when process preferences do matter.

---

[64] See for instance Osborne and Rubinstein (1994, p. 12).

### 3.4.1 Set-up

When an individual acts, he implements a *process* while expecting some *consequences*. The set $B_i$ of available behaviors is included in the Cartesian product of the set $A_i$ of processes of individual $i$ and the set $C_i$ of consequences. We denote $B_i \subset A_i \times C_i$, and $b_i = (a_i, g_i(a_i, a_{-i})) \in B_i$. For each individual, axioms 1 (weak ordering of behavioral preferences), axiom 2 (weak ordering of process preferences), axiom 3 (weak ordering of consequential preferences) and axiom 4 (optimality) introduced in chapter 1 are deemed to hold.

### 3.4.2 Rationality in Social games

We can define rational behavior in a game-theoretic context as follows[65].

**Definition 10** *A behavior* $(a_i^*, g_i(a_i^*, a_{-i}))$ *is* ***rational*** *if and only if, for all* $a_i \in A_i$ : $(a_i^*, g_i(a_i^*, a_{-i})) \succsim_i^B (a_i, g_i(a_i, a_{-i}))$.

Rational behavior is thus defined *dependently* of a set of processes for all other individuals (denoted $a_{-i}$).

### 3.4.3 A Nash Equilibrium Solution Concept Integrating Process Preferences

If we now characterize a profile of behaviors—one for each individual, such that all behaviors are rational, we come to the classical solution concept of a Nash Equilibrium. Such a solution can be written without difficulty in the case of a preference relation over behaviors instead of a preference relation over consequences only.

**Definition 11** *A profile of behaviors* $a^* = (a_i^*, a_{-i}^*)$ *is a* ***Nash Equilibrium*** *if and only if, for all* $i \in N$*, for all* $a_i \in A_i$ : $(a_i^*, g_i(a_i^*, a_{-i}^*)) \succsim_i^B (a_i, g_i(a_i, a_{-i}^*))$.

It may appear abusive to denote a profile of behaviors by a profile of processes $a^*$. This is not since a profile of processes specifies a process for each individual and thus a consequence for each individual. Therefore, a profile of processes determine a profile of behaviors. This constitutes a simpler notation.

### 3.4.4 Social Norms and Social Games

In order to disentangle the consequential game from the social game, we may first specify the role of process preferences. Any individual $j$ is assumed to have process preferences over his processes $a_j \in A_j$. The profile of such process preferences are deemed to constitute a *social norm.*

**Definition 12** *A profile of process preferences* $\alpha = \times_{j \in N} \succsim_j^A$ *is a* ***social norm****.*

Finally, a strategic social game can be summarized by the following definition.

---

[65] For notational simplification, we do not write $\succsim_i^{A_i}$ and $\succsim_i^{B_i}$but $\succsim_i^A$ and $\succsim_i^B$ .

**Definition 13** *A* ***social game*** $\Gamma_\alpha = \langle N, (A_i), g, (\succsim_i^C), \alpha \rangle$ *is defined by*:
*(i) A Consequential Game* $\Gamma$.
*(ii) A Social Norm* $\alpha$.

In this manner, each game of consequences is "embedded" in a procedural context. No rational solution for a social game can be identified without making assumptions about the social norm. On the other hand, observation of behavior serves to reveal such social norm. It is first proposed to refine the solution concept of equilibrium through the qualitative property of optimality.

### 3.4.5 An Optimal Nash Equilibrium that is Focal

Consider a situation where each individual has an available behavior whose process is preferred and whose consequence is the best of all equilibria. In this *optimal* case, there are "good reasons" to reach such an equilibrium and I propose to call such a "focal" or "salient" equilibrium an *Optimal Nash Equilibrium*[66]. Formally, denoting $A^*$ the set of profiles of behaviors that are Nash Equilibria, we can define an Optimal Nash Equilibrium as follows:

**Definition 14** *A profile of processes* $a^{**} = (a_i^{**}, a_{-i}^{**})$ *is an* ***Optimal Nash Equilibrium*** *if and only if,* $\forall\, i \in N : \forall a^* \in A^*, g_i(a^{**}) \succsim_i^C g_i(a^*)$ *and* $\forall a_i \in A_i, a_i^{**} \succsim_i^A a_i$.

An Optimal Nash Equilibrium may not exist. There are therefore strategic games without an Optimal Nash Equilibrium, even when there exists a unique Nash Equilibrium of behaviors[67]. However, when an Optimal Nash Equilibrium exists, it suffices to be composed of strict preferences to be unique.

**Corollary 4** *An Optimal Nash Equilibrium with strict process preferences or strict consequential preferences is unique.*

*Proof: Consider first the case of strict consequential preferences, i.e. for all individuals:* $g_i(a^{**}) \succ_i^C g_i(a)$ *for all* $a^* \in A^*$*, where* $a^{**}$ *is an Optimal Nash Equilibrium. If there was another Optimal Nash Equilibrium, say* $\grave{a}^{**} = (\grave{a}_i^{**}, \grave{a}_{-i}^{**})$*, then we would have* $\forall a^* \in A^*$, $g_i(\grave{a}^{**}) \succsim_i^C g_i(a^*)$*, in particular* $g_i(\grave{a}^{**}) \succsim_i^C g_i(a^{**})$ *which is a direct contradiction of the expression of strict consequential preferences. We now consider the case of strict process preferences, i.e.:* $\forall a_i \in A_i, a_i^{**} \succ_i^A a_i$*. If there was another Optimal Nash Equilibrium, say* $\grave{a}^{**}$*, then we would have* $\forall a_i \in A_i, \grave{a}_i^{**} \succsim_i^A a_i$ *and in particular* $\grave{a}_i^{**} \succsim_i^A a_i^{**}$ *which is a direct contradiction of the expression of strict process preferences* □

In this manner, the concept of Optimal Nash Equilibrium enables to select one single equilibrium among several ones. When it exists, it provides a sharp criterion to single out the "focal" or "salient" equilibrium in case of coordination issues.

---

[66] See in this respect the seminal works of Schelling (1960) and Lewis (1969). Another treatment of focal point is Sugden (1995).

[67] This is reflected by the condition on processes: $\forall a_i \in A_i, a_i^{**} \succsim_i^P a_i$ in the above definition.

A second corollary now links the concept of an Optimal Nash Equilibrium with the concept of a Pareto-efficient equilibrium. A Pareto-efficient equilibrium $\tilde{a}^*$ is the best possible equilibrium, i.e. for all $i$, for all $a_i^* \in A_i^* : (\tilde{a}_i^*, g_i(\tilde{a}_i^*, \tilde{a}_{-i}^*)) \succsim_i^B (a_i^*, g_i(a_i^*, a_{-i}^*))$.

**Corollary 5** *An Optimal Nash Equilibrium is always a Pareto-efficient equilibrium.*

*Proof: If a profile of processes $a^{**}$ is an Optimal Nash Equilibrium, then for all individuals, $a_i^{**} \succsim_i^A a_i$ and $g_i(a^{**}) \succsim_i^C g_i(a^*)$. We have in particular $a_i^{**} \succsim_i^A a_i^*$ and thus, by the optimality axiom, $(a_i^{**}, g_i(a^{**})) \succsim_i^B (a_i^*, g_i(a^*))$ for all $a_i^* \in A_i^* \square$*

An Optimal Nash Equilibrium thus sheds light on how individuals reach Pareto-efficient equilibria: individuals follow their preferred process and have no reason to deviate along that path since it guides them to the best possible equilibrium.

### 3.4.6 A Numerical Formulation of the Model

In the following section, a numerical formulation is presented so as to ease the treatment of concrete examples as well as their interpretation[68].

The sets $C_i$ of consequences are assumed to be mixture sets (see definition on page 36) and axioms 1 to 6 are assumed to hold. Consequential preferences of individuals are thus measurable by a *consequential utility function* $u_i$ that is a monotonic and linear functional form (a von Neumann and Morgenstern utility) on $C_i$. Individuals' process preferences are reflected as *weights* on consequential utility, i.e. outside the functional form. Although, these weights can be revealed as a ranking, they remain in general "unknown" to the observer[69]. In other words, preferences over behaviors are reflected through a *weighted von Neumann Morgenstern utility function*. We thus consider the model based on the following definition of rationality:

**Conjecture 6** *A behavior $(a^*, g(a_i^*, a_{-i}))$ is **rational** if and only if, for all $a_i \in A_i$ : $\alpha_i(a_i^*) \times u_i(g(a_i^*, a_{-i})) \geqslant \alpha_i(a_i) \times u_i(g(a_i, a_{-i}))$. Where $\alpha_i : a_i \longmapsto \alpha_i(a_i) > 0$ for all $a_i \in A_i$ is the **process utility function** of individual $i$. Where $u_i : c \longmapsto u_i(c) > 0$ for all $c \in C$ is the **consequential utility function** of individual $i$.*

The measurement of preferences is required to be strictly positive so that the property expressed by the optimality axiom (axiom 4 page 17) of the qualitative structure is appropriately reflected in the numerical formulation.

The origin of the interval scale $u$ measuring consequential preferences must be fixed such that $u_i(c_i) > 0$ for all $c_i \in C_i$. As a result, the ratio of process utilities is bounded by the ratio of consequential utilities as

---

[68] The manner by which this multiplicative form may be shown to be "necessary" seems to be an interesting problem.

[69] We may interpret them as particular type of "beliefs" that individuals have on themselves. This is interesting further research.

$$\frac{\alpha_i(a_i^*)}{\alpha_i(a_i)} \geqslant \frac{u_i(g(a_i, a_{-i}))}{u_i(g(a_i^*, a_{-i}))}. \tag{3.21}$$

The exact numerical value of the right hand side depends on the origin, say $\omega_0$ of the scale $u_i$. However, whether such ratio is greater or less than 1 does not depend on such origin (provided $u_i(c_i) > 0$ for all $c_i \in C_i$ still holds). Therefore, a ranking of process utilities nevertheless results.

This ordinal scale $\alpha$ "measuring" process preferences is normalized such that $\sum_{a_i \in A_i} \alpha_i(a_i) = 1$. If there are two processes $a_i^*$ and $a_i$, the limit-case $\alpha_i(a_i^*) = 1$ and $\alpha_i(a_i) = 0$ are interpreted as reflecting "*value-rational*" behavior[70]. Such behaviors are solely motivated by procedural judgments. In other words, the individual has a process utility on $a_i^*$ that is "so intense" that he implements $a_i^*$ whatever its consequence. This is always *relatively* to $a_i$ for which the individual then has a process disutility that is "so intense" that he won't implement $a_i$ whatever its consequence. These limit-cases are excluded through the two conditions $\alpha_i(a_i) > 0$ for all $a_i \in A_i$ and $\sum_{a_i \in A_i} \alpha_i(a_i) = 1$. The limit-case $\alpha_i(a_i^*) = \frac{1}{2}$ and $\alpha_i(a_i) = \frac{1}{2}$ corresponds to the neutrality of procedural judgments.

This numerical formulation helps to interpret the different possible empirical observations of rational behavior depending on process preferences.

The different social games that may result from different Prisoners' Dilemma of consequences are now studied.

## 3.5 The Influence of a Social Norm in the Prisoners' Dilemma

### 3.5.1 Set-up

Behaviors of individual $i$ are $(Co, g_i(Co, Co))$, $(Co, g_i(Co, De))$, $(De, g_i(De, Co))$, and $(De, g_i(De, De))$.

Individual $i$' consequential utility is given by:

$$u_i(g_i(De, Co)) = w_i \text{ } (i \text{ defects but} - i \text{ cooperates}), \tag{3.22}$$

$$u_i(g_i(Co, Co)) = x_i \text{ (both cooperate)}, \tag{3.23}$$

$$u_i(g_i(De, De)) = y_i \text{ (both defect)}. \tag{3.24}$$

$$u_i(g_i(Co, De)) = z_i \text{ } (i \text{ cooperates but} - i \text{ defects}), \tag{3.25}$$

We have $w_i$, $x_i$, $y_i$, and $z_i \in$ Re, as well as $w_i > x_i > y_i > z_i > 0$. The game of consequences can be represented in its normal form by the following matrix where processes of individual 1 are stated in rows and processes of individual 2 are stated in columns. In each box, consequential utility of individual 1 is stated first:

---

[70] This is the terminology of Weber (1922, p. 24).

*Table 3.1: The Prisoners' Dilemma of consequences*

| | Cooperation $(Co)$ | Defection $(De)$ |
|---|---|---|
| Cooperation $(Co)$ | $x_1, x_2$ | $z_1, w_2$ |
| Defection $(De)$ | $w_1, z_2$ | $y_1, y_2$ |

Individual $i$' process utility is reflected by:

$$\alpha_i : \begin{cases} \alpha_i(Co) = \alpha_i \\ \alpha_i(De) = 1 - \alpha_i \end{cases} \quad with \;\; 0 < \alpha_i < 1. \tag{3.26}$$

The profile $\alpha = (\alpha_1, \alpha_2)$constitutes a *social norm*.

We can now represent the social game combining consequential and process utility:

*Table 3.2*: *The Social Game*

| | Cooperation $(Co)$ | Defection $(De)$ |
|---|---|---|
| Cooperation $(Co)$ | $\alpha_1 \times x_1, \alpha_2 \times x_2$ | $\alpha_1 \times z_1, (1-\alpha_2) \times w_2$ |
| Defection $(De)$ | $(1-\alpha_1) \times w_1, \alpha_2 \times z_2$ | $(1-\alpha_1) \times y_1, (1-\alpha_2) \times y_2$ |

This social game is now analyzed in a standard manner.

### 3.5.2 Existence of Nash Equilibria

Applying the definition of a Nash Equilibrium of behaviors for the profile of processes $(Co, Co)$, cooperation for both individuals is a Nash Equilibrium of behaviors if and only if:

$$\Longleftrightarrow \forall\, i \in N,\; \alpha_i(Co) \times u_i(g_i(Co, Co)) \geqslant \alpha_i(De) \times u_i(g_i(De, Co)) \tag{3.27}$$

$$\Longleftrightarrow \alpha_1 x_1 \geqslant (1-\alpha_1) w_1 \;\; and \;\; \alpha_2 x_2 \geqslant (1-\alpha_2) w_2 \tag{3.28}$$

$$\Longleftrightarrow \alpha_1 \geqslant \frac{w_1}{w_1 + x_1} \;\; and \;\; \alpha_2 \geqslant \frac{w_2}{w_2 + x_2}. \tag{3.29}$$

Similarly, $(De, De)$ is a Nash Equilibrium of behaviors if and only if:

$$\alpha_1 \leqslant \frac{y_1}{y_1 + z_1} \;\; and \;\; \alpha_2 \leqslant \frac{y_2}{y_2 + z_2}. \tag{3.30}$$

This game may also include *asymmetric* equilibria. Individual $i$ taking process $Co$ and individual $-i$ taking process $De$, that is profile of processes $(Co, De)$ is a Nash Equilibrium of behaviors if and only if, for $i = 1, 2$ :

$$\alpha_i \geqslant \frac{y_i}{y_i + z_i} \;\; and \;\; \alpha_{-i} \leqslant \frac{w_{-i}}{w_{-i} + x_{-i}}. \tag{3.31}$$

Having determined the conditions of existence of Nash Equilibria, we characterize now situations where several of them coexist.

### 3.5.3 Selecting the Optimal Nash Equilibrium among Multiple Nash Equilibria

Firstly, we can see that asymmetric equilibria cannot coexist with symmetric equilibria, conditions 3.31 being not compatible with either 3.29 or 3.30.

Secondly, conditions for the coexistence of symmetric Nash Equilibria are:

$$\frac{y_i}{y_i + z_i} \geqslant \alpha_i \geqslant \frac{w_i}{w_i + x_i}, \forall i \in N. \tag{3.32}$$

When it exists, equilibrium $(Co, Co)$ is always the unique Optimal Nash Equilibrium. This is because of the construction $\frac{w_i}{w_i+x_i} > \frac{1}{2}$, $\forall\, i \in N$. Therefore, $\alpha_i \geqslant \frac{w_i}{w_i+x_i}, \forall i \implies \alpha_i \geqslant \frac{1}{2}$ and thus $\alpha_i(Co) > \alpha_i(De)$. The equilibrium $(De, De)$ is the unique Optimal Nash Equilibrium if and only if $\alpha_i \leqslant \frac{1}{2}, \forall\, i \in N$.

Thirdly, conditions for the coexistence of asymmetric Nash Equilibria are:

$$\frac{w_i}{w_i + y_i} \geqslant \alpha_i \geqslant \frac{y_i}{y_i + z_i}, \forall i \in N. \tag{3.33}$$

Finally, a special case arises where no Nash Equilibrium exists. The conditions for absence of Nash Equilibrium are (pairs of indices $(i, -i)$ being $(1, 2)$ or $(2, 1)$):

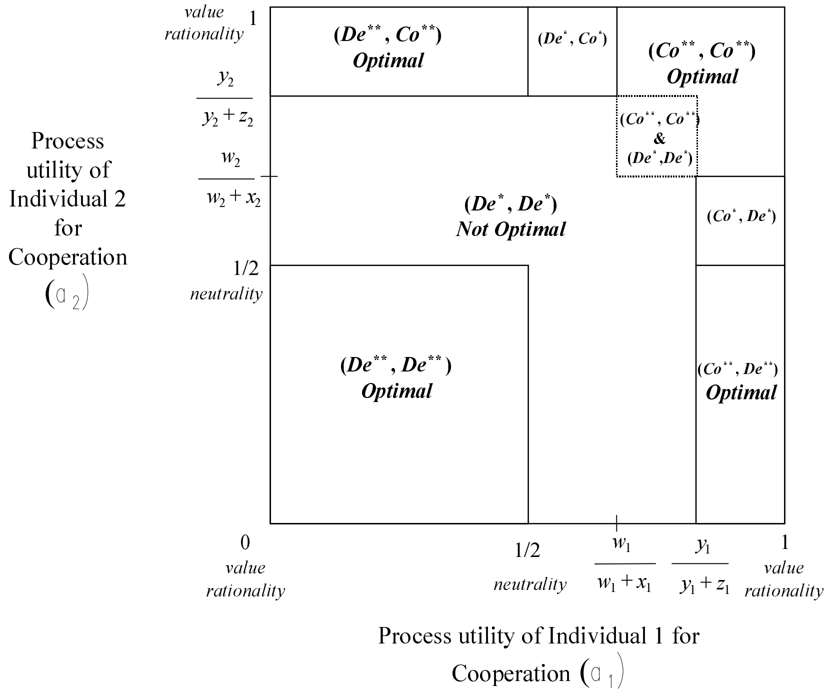
$$\frac{y_i}{y_i + z_i} \geqslant \alpha_i \geqslant \frac{w_i}{w_i + x_i} \quad and \quad \frac{w_{-i}}{w_{-i} + x_{-i}} \geqslant \alpha_{-i} \geqslant \frac{y_{-i}}{y_{-i} + z_{-i}}. \tag{3.34}$$

### 3.5.4 Graphical Summary of Solutions

The interest of the approach is not to determine a unique rational solution for the Prisoners' Dilemma but to make use of empirical observation of rational behavior to reveal process preferences. We express rational solutions of the social game depending on process preferences. To this purpose, we represent graphically the different Nash Equilibria using the process utilities $\alpha_1$ and $\alpha_2$ of individuals 1 and 2 as coordinates. These process utilities have a range from 0 to 1 excluded. Process utilities of 0 and 1 are *value-rational* and correspond to the limit-case where individuals are solely concerned by the process of their behavior. Process utilities of $1/2$ are *neutral* and correspond to the limit-case where individuals are solely concerned by the consequence of their behavior. We distinguish three cases corresponding to three classes of Prisoners' Dilemma games. In all of them, observation of a cooperating behavior reveals a process utility for cooperating while observation of defection does not reveal whether the individual has a process utility for defecting or whether he has a process utility for cooperation, although too weak to influence his behavior.

Case 1: $\frac{w_i}{w_i+x_i} \leqslant \frac{y_i}{y_i+z_i}$ for all individuals

As shown in Figure 13, a Nash Equilibrium always exists in this case. Symmetric equilibria of Cooperation and Defection can coexist although *only the cooperation equilibrium is optimal* (an Optimal Nash Equilibrium is depicted with the



13.Nash Equilibria in a Prisoners' Dilemma: Case 1

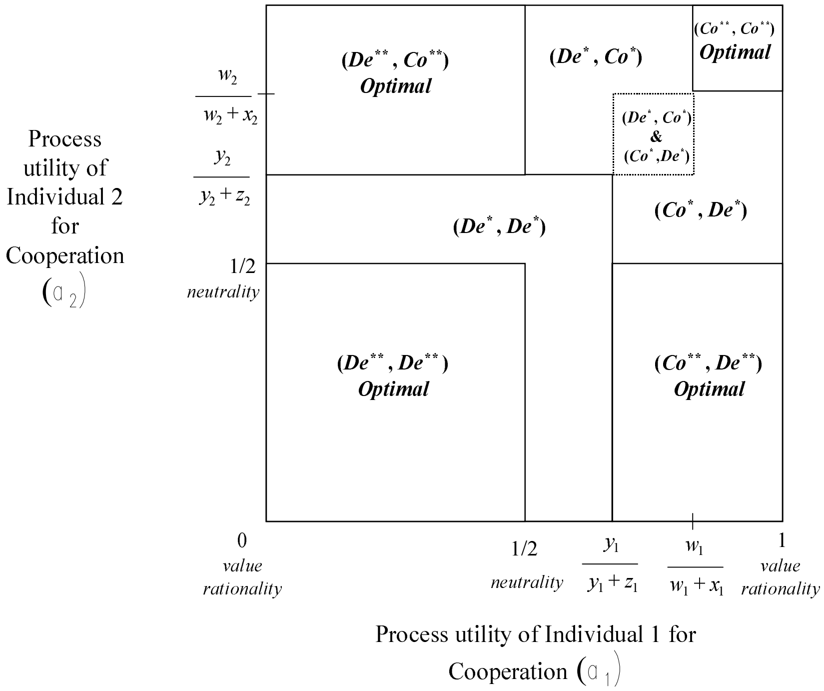
double star). At each asymmetric corner, there exists an asymmetric Nash Equilibrium which is Optimal for both individuals: one "sacrifices" himself for the other. The Defection equilibrium is Optimal only when both individuals have a process utility for Defection.

Case 2: $\frac{w_i}{w_i+x_i} \geqslant \frac{y_i}{y_i+z_i}$ for all individuals

As shown in Figure 14, a Nash Equilibrium also always exists in this case. The symmetric and optimal equilibrium of cooperation is restricted to stronger process preferences than in case 1. The symmetric equilibrium of defection holds for symmetrical procedural preferences close to neutrality but it never coexists with the cooperation equilibrium. At each asymmetric corner, there exists asymmetric Optimal Nash Equilibria similarly to case 1. These equilibria may also coexist although none is optimal.

Cases 3: $\frac{y_i}{y_i+z_i} \geqslant \frac{w_i}{w_i+x_i}$ and $\frac{w_{-i}}{w_{-i}+x_{-i}} \geqslant \frac{y_{-i}}{y_{-i}+z_{-i}}$

Figure 15 indeed represents two permutable cases depending whether we take $i = 1 (-i = 2)$ or $i = 2$ $(-i = 1)$. These third cases are characterized by their asymmetry, which entails the absence of equilibrium for some values of $(\alpha_1, \alpha_2)$. There is no coexistence of equilibria.



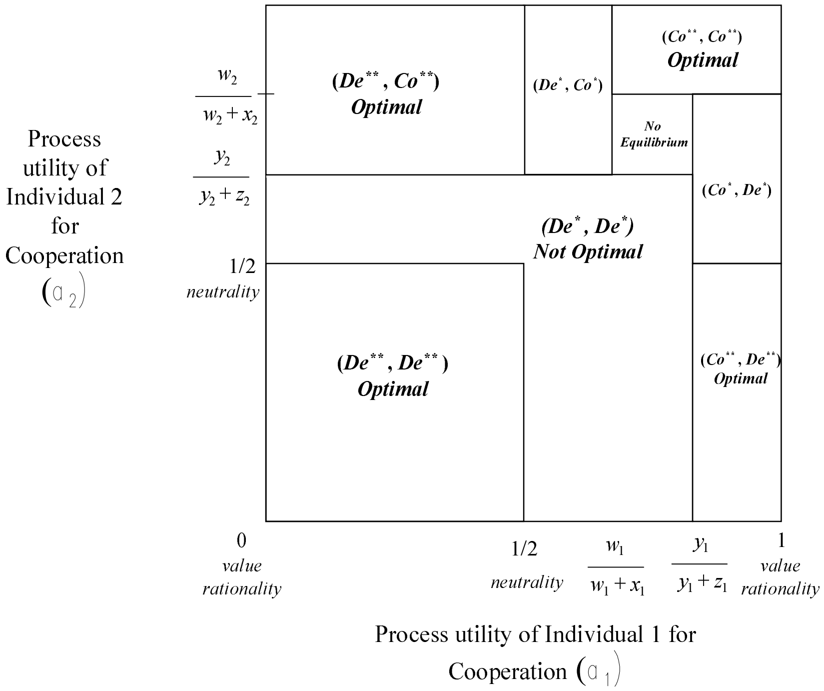
14.Nash Equilibria in a Prisoners' Dilemma: Case 2

## 3.6 Conclusion to Chapter 3

This chapter attempts to extend game theory such as to explicitly integrate procedural and consequential preferences. This allows to state the solution concept of Nash Equilibrium with preferences for behaviors instead of preferences for consequences in a straightforward manner.

The structure of our approach leads to the introduction of the concept of *Optimal Nash Equilibrium* which is a refinement of the Nash Equilibrium concept. Such a selection characterizes a profile of behaviors composed of procedurally preferred processes and preferred consequences at equilibrium. There are "good reasons" to reach such equilibrium since all individuals prefer the process to attain it, and also prefer the consequences they reach over all consequences reached at equilibrium. When such an optimal equilibrium exists, and either process or consequential preferences are strict, then the Optimal Nash Equilibrium is unique. Moreover, an Optimal Nash Equilibrium is always Pareto-efficient. By justifying the existence of a "focal equilibrium", the integration of process preferences in the definition of rationality may lead to a richer understanding of the general notion of equilibrium.

Empirically, this approach sheds light on the distinction between the *consequential game* and the *social game* actually perceived and acted by the individuals. The social game does integrate preferences over processes *per se*. Depending on procedural preferences, there exists many solutions to such games, even when preferences for consequences have been "accurately measured" beforehand. In particular, the ob-



15.Nash Equilibria in a Prisoners' Dilemma: Cases 3

servation of cooperation in the Prisoners' Dilemma does not reveal "irrationality", but reveals *the influence of a social norm on consequential behavior*. Sufficiently strong process preferences with regards to the consequences at stake lead to both individuals cooperating as the unique Optimal Nash Equilibrium of the game. In these situations, consequential behavior is influenced by a social norm that helps the collectivity of individuals to reach a more efficient state.

Why a particular individual does have procedural preferences or is neutral remains on open question, which is a matter of explaining preferences, not behavior. One may argue that process preferences may come from tradition, from membership of a social community, from cognitive consistency, etc. I have treated them as given. It may be possible to model their genesis more precisely. One possible way to proceed is to model how pre-play communication induces process preferences for not lying. Then, rational communication may be conveniently modelled so as to explore how individuals choose the social norms that can guide them to an Optimal Nash Equilibrium.

# Conclusion

> *"Individuals are the only objects of which numbers cannot be significantly asserted".*
> *B. Russell*

If the standard rationality hypothesis is that "behavior can be represented as the maximization of a suitably restricted utility function", then the model explored in this dissertation may be non-standard[71]. The consideration of entities of the type (process, consequence) seems to provide intuitive description of rational behavior, although the formalism implied by such postulate has been difficult to circumvent with all the desirable precision and clarity. The representation of rational behavior by a suitably restricted utility function—when behavior is defined as couples of processes and their consequences—seems indeed problematic. This conclusion does not intend to provide answers to the questions that have been left aside along the dissertation. I rather attempt to share a view on the general problem raised and on its interpretation with regards to scientific reasoning.

It seems to me that a fundamental difficulty encountered in this dissertation can be traced back to the notion of singularity. By singularity, I mean the specificity pertaining to each individual who performs an act of choice. This singularity is abstracted from the methodology that consists in representing preferences by a function because it is not a property of the entities over which preferences apply. For instance, that an individual chooses *this* pear rather than *this* apple is not a property of *this* pear nor of *this* apple. "Naturally", we can represent *this pear* by *the class of pears* and *this apple* by *the class of apples*. Then, it is possible to express that pears are preferred to apples as a property of the class of pears *or* as a property of the class of apples, opening the possibility of a functional representation and thus of measurement. However, it may not adequately reflect that in a particular case, the individual *chooses to eat this* apple he received from his grand-mother rather than *this* pear he bought at the supermarket. In such a situation, equivalence classes have to be re-defined appropriately. The problem is that such re-definition may occur each time the model is related to experience[72]. Once fruits can be combined with probabilities, preferences become quantitative and it is meaningful to express whether pears are "more preferred" to apples than bananas are to oranges. This does not however constitute a solution to the above issue. In real life, part of the act of choice is always and unavoidably singular.

We may reject the methodology of measurement under the critics that it excludes such singularity. I have argued for another approach, building on measurement rather than invalidating it.

On one hand, consequences are defined as *classes of consequences*, like in the von Neumann and Morgenstern approach. In this manner, they can be measured

---

[71] See the beginning of Chapter 1.

[72] Recall the quote from Stigler page 33.

with appropriate axioms. On the other hand, consequences are *related* to individuals through the notion of processes. In this manner, each process is *singular*. From there, our approach has taken two directions. First, leaving processes singular and supposing that preferences over such entities may exist (process preferences). The implication being that process preferences can be revealed in a very intuitive manner although the logic is quite peculiar and should be clarified. Second, making explicit the representation of processes through the choice of process attributes. This second direction, when combined with the first, has been argued to constrain the interpretive framework so that explanation of behavior through attributes over processes must follow a logical framework.

This introduction of the singularity of the process of behavior combined with an extensive definition of consequences lies at the heart of our model. Along the dissertation, it has been given the interpretation that the singularity comes from individuals themselves, their own values, their own adaptation to the context, their experience, etc. Following the analogy with mechanics introduced in the first chapter, a comparison with the standard model of the representational theory of measurement—the equally-armed balance—provides intuition to this interpretation.

Consider that an individual acts as if he were a balance. The behavior of the individual, as the behavior of the balance, are the observable entities. If the individual chooses one behavior and not another, then a relation between the two behaviors can be expressed in terms of preferences. A similar relation between each side of the balance can be expressed when objects are placed on its pans. Further assumptions can be made so that these comparisons become measurable. For the individual, it is the combination of consequences with probabilities. For the balance, it is an homogeneous substance that can be "concatenated" so that each object on a pan of the balance can be balanced with a precise quantity of such "standard units" placed on the other side. Such a quantity then characterizes a whole equivalence class of objects: the ones that have the same mass. As a result, behavior of the balance can be formulated as determined by masses. In the analogy, behavior of the individual can be formulated as determined by probabilistic consequences. However, while there is a large empirical evidence supporting the formulation in the case of the balance, there is a large empirical evidence invalidating the formulation in the case of the individual.

One distinction in the analogy is that on one hand the balance is a measuring device that is material, and on the other hand, individuals are themselves the measuring devices. It is classical in mechanics to suppose that the measuring device is known and invariant. This is indeed the assumption that allows to construct classes of observations that are equivalent in the sense that they are supposed to derive from the same measuring device. This model may be said to relax the assumption that any measuring device must be *invariant*.

In other words, considering that individuals have values is the analog of considering that the fulcrum of the balance may not be at the middle of the arm. This corresponds to a bias in the behavior of individuals, as it does for the behavior of the balance. Values of individual may change and adapt to the context and to experience, just as the fulcrum of our peculiar balance may move along the arm. We may think that nothing interesting would be left and indeed, in the absence of a meta-measuring device, changes in the position of the fulcrum cannot be known with precision. Nevertheless, maintaining the concept of “masses” ensures a qualitative comparison of the different positions of the fulcrum. Some usefulness for formal models thus remains so as to study measuring devices. This way to “inverse” scientific reasoning may reveal some interesting features. In the study of rational behavior, it consists in building of consequential models so as to better understand processes, i.e., individuals.

# References

Allais, M.: 1953, “Le comportement de l’homme rationnel devant le risque: critique des postulats et axiomes de l’école américaine”, *Econometrica,* **21**, 503–546.

Anscombe, F.J., and R.J. Aumann: 1963, “A Definition of Subjective Probability”, *Annals of Mathematical Statistics,* **34**, 199-205.

Arrow, K. J.: 1994, “Methodological Individualism and Social Knowledge”, *The American Economic Review*, **84**, 1-9.

Aumann, R. J.: 1997, “Rationality and Bounded Rationality”, *Games and Economic Behavior*, **21**, 2-14.

Aumann, R. J.: 1998, “Common Priors: A reply to Gul”, *Econometrica,* **66**, 929–938.

Axelrod, R.: 1984, *The Evolution of Cooperation*, Basic Books, New York.

Barbera, S., Hammond, P.J., Seidl, C. (eds): 1998, *Handbook of Utility Theory: Volume 1 Principles*, Kluwer Academic Press, Boston.

Beard, T. R., Beil, R.O.: 1994, “Do People Rely on the Self-interested Maximization of Others ? An Experimental Test”, *Management Science*, **40,** 252-262.

Bell, D., H. Raiffa, and A. Tversky (ed.): 1988, *Decision Making: Descriptive, normative, and prescriptive interactions*, Cambridge University Press, Cambridge.

Bennett, P., Howard, N.: 1996, “Rationality, emotion and preference change: Drama-theoretic models of choice”, *European Journal of Operational Research*, **92,** 603-614.

Bernoulli, Daniel: 1954, “Exposition of a new theory on the measurement of risk”, *Econometrica*, **22**, 23-26.

Binmore, K.: 1994, *Game theory and the Social Contract - Volume I*: *Playing Fair*, MIT Press, Cambridge MA.

Boudon, R.: 1987, “The Individualistic Tradition in Sociology”, in Jeffrey C.A., B. Giesen, R. Münch, N. Smelser (eds), *The Micro-Macro Link,* University of California Press, Berkeley, 45-70.

Boudon, R.: 1994, *The Art of Self-Persuasion*, Polity Press, London.

Boudon, R.: 1996, “The ‘Cognitivist Model’ - A Generalized ‘Rational-Choice’ Model”, *Rationality and Society,* **8**, 123-50.

Boudon, R.: 1997, “Le ’paradoxe du vote’ et la théorie de la rationalité”, *Revue Française de Sociologie*, **38**, 217-227.

Boudon, R.: 1998, “Limitations of Rational Choice Theory”, *American Journal of Sociology,* **104**, 817-828.

Brennan, G., and A. Hamlin: 1996, “Two Theories of Rational Voting”, *Discussion Papers in Economics and Econometrics No. 9632*, ISSN 0966-4246.

Camerer, C. and M. Weber: 1993, “Recent Developments in Modelling Preferences: Uncertainty and Ambiguity”, *Journal of Risk and Uncertainty*, **5**, 325-370.

Clyman, Dana R.: 1995, "Unreasonable Rationality ?", *Management Science*, **41,** 1538-1548.

Conlisk, J.: 1996, "Why Bounded Rationality ?", *Journal of Economic Literature*, **34**, 669-700.

de Finetti, B.: 1937, "La prévision : ses lois logiques, ses sources subjective"*, Annales de l'Institut Henri Poincaré,* **VII**, 1-68.

de Finetti, B.: 1974, *Theory of Probability,* Volumes 1 & 2, John Wiley and Sons, Chichester, England.

Donaldson, C. and P. Shackley: 1997, "Does 'Process Utility' exist? A case study of willingness to pay for laparoscopic cholecystectomy", *Social Science Medicine*, **44**, 669-707.

Dreier, J.: 1996, "Rational Preference: Decision Theory as a Theory of Practical Rationality", *Theory and Decision*, **40,** 249-276.

Drèze, J.H.: 1958, *Individual Decision Making Under Partially Controllable Uncertainty,* Unpublished Ph.D. thesis, Colombia University.

Drèze, J.H.: 1987, *Essays on economic decisions under uncertainty,* Cambridge University Press, Cambridge.

Ellsberg, D.: 1961, "Risk, Ambiguity and the Savage axioms", *Quarterly Journal of Economics*, **75**, 643-669.

Elster, J.: 1979, *Ulysses and the Sirens*, Cambridge University Press, Cambridge MA.

Elster, J.: 1989a, *Nuts and Bolts for the Social Sciences*, Cambridge University Press, Cambridge MA.

Elster, J.: 1989b, "Social Norms and Economic Theory", *Journal of Economic Perspectives*, **3**, 99-117.

Festinger, L.: 1957, *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford.

Fillieule, R.: 1996, "Frames, Inferences and Rationality", *Rationality and Society,* **8**, 151-165.

Fishburn, Peter C.: 1970, *Utility Theory for Decision Making*, John Wiley & Sons, New York.

Fishburn, P. C.: 1973, "A Mixture-Set Axiomatization of Conditional Subjective Expected Utility", *Econometrica*, **41**, 1-25.

Fishburn, P. C.: 1980, "A Simple Model for the Utility of Gambling", *Psychometrika*, **45**, 435–448.

Fishburn, Peter C.: 1981, "Subjective Expected Utility: A review of Normative Theories", *Theory and Decision*, **13**, 139-199.

Fishburn, P. C.: 1982, *The Foundations of Expected Utility,* D. Reidel, Dordrecht, Holland.

Fishburn, P. C.: 1988a, *Nonlinear Preference and Utility Theory,* The Johns Hopkins University, Baltimore and London.

Fishburn, Peter C.: 1988b, “Expected Utility: An Anniversay and a New Era”, *Journal of Risk and Uncertainty*, **1,** 267-283.

Fishburn, P. C.: 1989a, “Retrospective on the Utility Theory of von Neumann and Morgenstern”, *Journal of Risk and Uncertainty*, **2**, 127–158.

Fishburn, Peter C.: 1989b, “Foundations of Decision Analysis: Along the Way”, *Management Science*, Vol. 35, April 1989, pp. 387-405.

Fishburn, P. C. and P. Wakker: 1995, “The Invention of the Independence Condition for Preferences”, *Management Science*, **41**, 1130–1144.

Frank, R.: 1988, *Passions within Reason*, Norton, New York.

Fudenberg, D. and D. K. Levine: 1998, *The theory of learning in games*, MIT Press, Cambridge MA..

Gärling, T., K. Axhausen and M. Brydsten: 1996, “Travel Choice and the Goal/Process Utility Distinction”, *Applied Cognitive Psychology*, **10**, 65-74.

Gauthier, D.: 1986, *Morals by Agreement*, Clarendon Press, Oxford.

Geanakoplos, J. and D. Pearce: 1989, “Psychological Games and Sequential Rationality”, *Games and Economic Behavior*, **1**, 60-79.

Gilboa, I., and Schmeidler, D.: 1988, “Information Dependent Games: Can Common Sense be Common Knowledge?”, *Economic Letters*, **27**, 215-221.

Granovetter, M.: 1985, “Economic Action and Social Structure: The Problem of Embeddedness”, *American Journal of Sociology*, **91**, 481-510.

Grize, J.B.: 1994, “Pensée logico-mathématique et sémiologie du langage”, in *Pensée Logico-Mathématique*, ed. by O. Houdé et D. Miéville, P.U.F., Paris, 121–204.

Hammond, P.J.: 1988, “Consequentialist Foundations of Expected Utility”, *Theory and Decision*, Vol. 25, 25-78.

Hammond, P. J.: 1996, “Consequentialist Decision Theory and Utilitarian Ethics”, in *Ethics, Rationality, and Economic Behavior*, Farina, F., Hahn, F. and S. Vannucci (eds), Clarendon Press, Oxford.

Hammond, P.J.: 1998, “Objective Expected Utility”, in *Handbook of Utility Theory: Volume 1 Principles*, Barbera, S., Hammond, P.J., Seidl, C. (eds), Kluwer Academic Press, Boston, 143-211.

Harsanyi, J. C.: 1978, “Bayesian Decision Theory and Utilitarian Ethics”, *American Economic Review*, **68**, 223–228.

Hechter, M.: 1994, “The Role of Values in Rational Choice Theory”, *Rationality and Society*, **6**, 318-333.

Herstein, I. N. and J. Milnor: 1953, “An Axiomatic Approach to Measurable Utility”, *Econometrica*, **21**, 291–297.

Kadane, J.B., and R.L. Winkler: 1988, “Separating Probability Elicitation from Utilities”, *Journal of American Statistical Association*, **83**, 357-363.

Kahneman, D. and D. T. Miller: 1986, “Norm Theory: Comparing Reality to Its Alternatives”, *Psychological Review*, **93**, 136-153.

Kahneman, D. and A. Tversky: 1979, “Prospect Theory: an Analysis of Decisions under Risk”, *Econometrica*, **47**, 263–291.

Kahneman, D. and A. Tversky: 1984, “Choices, Values, and Frames”, *American Psychologist*, **39**, 341-350.

Kahneman, D., Wakker, P., and R. Sarin: 1997, “Back to Bentham? Explorations of Experienced Utility”, *The Quarterly Journal of Economics*, **112**, 375–406.

Kandori, M.: 1992, “Social Norms and Community Enforcement”, *Review of Economic Studies*, **59**, 63-80.

Keeney, R. L. and H. Raiffa: 1976, *Decisions with Multiple Objectives*: *Preferences and Value Tradeoffs,* Wiley, New York.

Krantz, D.H., Luce, R.D., Suppes, P., and A. Tversky: 1971, *Foundations of Measurement*, Vol. 1, Academic press, New York and London.

Kreps, D. M.: 1988, *Notes on the Theory of Choice*, Westview Press, Boulder and London.

Laibson, D. and R. Zeckhauser: 1998, “Amos Tversky and the Ascent of Behavioral Economics”, *Journal of Risk and Uncertainty*, **16**, 7-47.

Levy-Garboua: 1998, “Expected Utility and Cognitive Consistency”, *Working-Paper*, LAMIA, Université Paris I Sorbonne, Paris.

Lewis, D. K.: 1969, *Convention*: *A Philosophical Study*, Harvard University Press, Cambridge MA.

Lewis, D. K.: 1973, *Counterfactuals*, Basil Blackwell, Oxford UK.

Lind, A.E. and Tyler, T.R.: 1988, *The Social Psychology of Procedural Justice*, Plenum Press, New York.

Luce, R. D.: 1992, “Where Does Subjective Expected Utility Fail Descriptively?”*, Journal of Risk and Uncertainty,* **5**, 5-27.

Luce, R. D.: 1996, “The Ongoing Dialog between Empirical Science and Measurement Theory”*, Journal of Mathematical Psychology,* **40**, 78-98.

Luce, R. D. and L. Narens: 1985, “Classification of Concatenation Structures according to Scale Type”, *Journal of Mathematical Psychology*, **29**, 9-72.

Luce, R. D. and L. Narens: 1987, “Measurement Scales on the Continuum”*, Science*, **236**, 1527-1532.

Luce, R. D. and H. Raiffa: 1957, *Games and Decisions,* Wiley, New York.

Luce, R. D. and D. von Winterfeld: 1994, “What Common Ground Exists for Descriptive, Prescriptive, and Normative Utility Theories?”*, Management Science*, **40**, 263-279.

Machina, M. J.: 1982, “‘Expected Utility’ Analysis without the Independence Axiom”, *Econometrica*, **50**, 1069–1079.

Machina, M. J.: 1987a, “Decision-Making in the Presence of Risk”, *Science*, **236**, 537–543.

Machina, Mark J.: 1987b, “Choice Under Uncertainty: Problems Solved and Unsolved”, *Economic Perspectives*, **1**.

Malinvaud, E.: 1952, “Note on the von Neumann-Morgenstern’s Strong Independence Axiom”, *Econometrica*, **20,** 679.

March: 1988, “Bounded Rationality, Ambiguity, and the Engineering of Choice”, in *Decision Making: Descriptive, normative, and prescriptive interactions*, ed. by D. Bell, H. Raiffa, and A. Tversky, Cambridge University Press, Cambridge, 33-57.

Marschak, J.: 1950, “Uncertain Prospects, and Measurable Utility”, *Econometrica*, **18**, 111–141.

Mooney, G.: 1994, *Key Issues in Health Economics*, Harvester Wheatsheaf, London.

Morgenstern, O.: 1979, “Some reflections on Utility”, in *Expected Utility Hypothesis and the Allais Paradox: Contemporary Discussions of Decisions Under Uncertainty with Allais’ Rejoinder*, ed. by M. Allais and O. Hagen, D. Reidel, Dordrecht, Holland, 175–183.

Mongin, P., and C. d’Aspremont: 1998, “Utility Theory and Ethics”, in *Handbook of Utility Theory: Volume 1 Principles*, Barbera, S., Hammond, P.J., Seidl, C. (eds), Kluwer Academic Press, Boston, 371-481.

Nau, R. F. and McCardle, K. F.: 1991, “Arbitrage, Rationality, and Equilibrium”, *Theory and Decision*, **31,** 199-240.

Nozick, R.: 1993, *The Nature of Rationality*, Princeton University Press, Princeton.

Osborne, M. J., and Rubinstein A.: 1994, *A Course in Game Theory*, MIT Press, Cambridge MA.

Osborne, M. J. and Rubinstein A.: 1998, “Games with Procedurally Rational Players”, *The American Economic Review, **88**,* 834-847.

Page, A.N.: 1968, *Utility Theory: A Book of readings,* Wiley, New York.

Pareto, V.: 1906, *Manuale di Economia politica,* Societa Editrice Libraria, Milan. Parts of the second edition (1927) are reprinted in *Utility Theory: A Book of readings*, Alfred N. Page, 1968, Wiley, New York, 108-181.

Pascal, B.: 1670, *Pensées,* Port-Royal, Paris.

Petersen, T.: 1992a, “Individual, Collective, and Systems Rationality in Work Groups : Dilemmas and Non-market Solutions”, *Rationality and Society*, **4**, 332-355.

Petersen, T.: 1992b, “Individual, Collective, and Systems Rationality in Work Groups : Dilemmas and Market-Type Solutions”, *American Journal of Sociology*, **98**, 469-510.

Piaget, J. (ed.): 1967, *Logique et Connaissance Scientifique*, Gallimard, Paris.

Pope, R. E.: 1995, “Towards a More Precise Decision Framework, A Separation of Negative Utility of Chance from Diminishing Marginal Utility and the Preference for Safety”, *Theory and Decision*, **39**, 241–265.

Pope, R. E.: 1998, “Attractions to and Repulsions from Chance”, in *Game Theory, Experience, Rationality,* ed. by W. Leinfellner and E. Köhler, Kluwer Academic Publishers, 95–107.

Rabin, M.: 1993, “Incorporating Fairness into Game Theory and Economics”, *American Economic Review*, **83**, 1281-1302.

Ramsey, F. P.: 1931, “Truth and Probability”, *in The Foundations of Mathematics and Other Logical Essays*, 156-198, Routledge and Kegan Paul, London.

Rapoport, A., & Chammah, A.: 1965, *Prisoner’s Dilemma*, University of Michigan press, Ann Arbor.

Rapoport, A., Guyer, M.J., & Gordon, D.G.: 1976, *The 2 x 2 game*, University of Michigan Press, Ann Arbor.

Rapoport, A.: 1988, “Experiments with n-person social traps I: Prisoner’s Dilemma, weak Prisoner’s Dilemma, Volunteer’s Dilemma, and Largest Number”, *Journal of Conflict Resolution*, **32**, 457-472.

Requier-Desjardins, D.: 1995, “Temps, Rationalité, Règles”, *Working Paper*, C3ED, Université de Versailles Saint-Quentin-en-Yvelines, France.

Rubinstein, A.: 1991, “Comments on the Interpretation of Game Theory”, *Econometrica*, **59**, 909-924.

Rubinstein, A.: 1998, *Modeling Bounded Rationality*, The MIT Press, Cambridge MA.

Sainfort. F., J.M. Deichtmann: 1996, “Decomposition of Utility Functions on Subsets of Product Sets”, *Operations Research*, **44**, 609-616.

Sally, D.: 1995, “Conversation and Cooperation in Social Dilemmas - A Meta-Analysis of Experiments from 1958 to 1992”, *Rationality and Society*, **7**, 58-92.

Sarin, R., Wakker, P.: 1994, “Folding Back in Decision Tree Analysis”, *Management Science*, **40**, 625-628.

Savage, L.J.: 1954, *The Foundations of Statistics*, Wiley Publications in Statistics, New York.

Schwager, K. W.: 1991, “The Representational Theory of Measurement: An Assessment”, *Psychological Bulletin*, **110**, 618-626.

Schelling, T. C.: 1960, *The strategy of Conflict*, Harvard University, Cambridge MA. Second edition 1980.

Schervish, M.J., T. Seidenfeld, and J.B. Kadane: 1990, “State-Dependent Utilities”, *Journal of American Statistical Association*, *85*, 840-847.

Schipper, F.: 1996, “Rationality and the Philosophy of Organization”, *Organization*, 3 (2) : 267-289.

Schoemaker, P. J.H..: 1982, “The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations”, *Journal of Economic Literature*, **20**, 529-563.

Sen, A.: 1986, “Foundations of social choice theory : an epilogue”, in *Foundations of social choice theory*, J. Elster and A. Hylland (eds), Cambridge University Press, Cambridge.

Sen, A.: 1995, “Rationality and Social Choice”, *The American Economic Review*, **85**, 1–24.

Sen, A.: 1997, “Maximization and the Act of Choice”, *Econometrica*, **65**, 745-779.

Shafir, E., and A. Tversky: 1992, “Thinking through uncertainty: Nonconsequential reasoning and choice”, *Cognitive Psychology,* **24**, 449-474.

Simon, H.A.: 1947, *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, The Free Press, New York, Fourth Edition, 1997.

Simon, H.A.: 1955, “A Behavioral Model of Rational Choice”, *Quarterly Journal of Economics*, **69**, 99-118.

Simon, H.A.: 1956, “Rational Choice and the Structure of the Environment”, *Psychological Review*, **63**, 129–138.

Simon, H.A.: 1976, “From substantive to procedural rationality”, in *Methods and Appraisal in Economics*, S. J. Latsis(ed.), Cambridge University Press, Cambridge MA.

Simon, H.A.: 1978, “Rationality as Process and as Product of Thought”, *The American Economic Review*, **68**, 1-16.

Simon, H.A.: 1979, “Rational Decision Making in Business Organizations”, *The American Economic Review*, **69,** 493-513.

Skiadas, C.: 1997, “Conditioning and Aggregation of Preferences”, *Econometrica*, **65**, 347-367.

Smelse, N.J.: 1992, “The Rational Choice Perspective : A Theoretical Assessment”, *Rationality and Society*, **4**, 381-410.

Stevens, S.S.: 1946, “On the theory of scales of measurement”, *Science*, **103**, 122-136.

Stigler, G. J..: 1950, “The development of utility theory: I, II”, *Journal of Political Economy*, **58**, 307-327, 373-396.

Sugden, R.: 1995, “A Theory of Focal Points”, *The Economic Journal*, **105,** 533-550.

Quine, W.V.: 1975, “On Empirically Equivalent Systems of the World”, *Erkenntnis*, **9**, 313-328.

Tversky, A., and D. Kahneman: 1974, “Judgment under Uncertainty : Heuristics and Biases”, *Science*, **185**, 1124-1131.

Tversky, A., and Kahneman, D.: 1981, “The framing of decisions and the psychology of choice”, *Science,* **211**, 453-458.

Tversky, A., and D. Kahneman: 1988, “Rational choice and the framing of decisions”, in *Decision Making: Descriptive, normative, and prescriptive interactions*, ed. by D. Bell, H. Raiffa, and A. Tversky, Cambridge University Press, Cambridge, 167-192.

Tversky, A. and E. Shafir: 1992, “Choice under Conflict : The Dynamics of Deferred Decision”, *Psychological Science*, **3**.

von Neumann, J.: 1951, “The General and Logical Theory of Automata”, in *Cerebral Mechanisms in Behavior: The Hixon Symposium*, Lloyd A. Jeffries (ed.), John Wiley and Sons, New York.

von Neumann, J., and O. Morgenstern: 1944, *Theory of Games and Economic Behavior,* Princeton University Press, Princeton, NJ, Second edition 1947, Third edition 1953.

Vriend, N.J..: 1996, “Rational Behavior and Economic Theory*”, Journal of Economic Behavior and Organization*, **29**, 263-285.

Wakker, P., and D. Deneffe: 1996, “Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown”, *Management Science*, **42**, 1131–1150.

Walliser, B.: 1989, *“*Instrumental Rationality and Cognitive Rationality*”, Theory and Decision,* **27**, 7-36.

Weber, M.: 1978, *Economy and Society: An Outline of Interpretive Sociology*, University of California Press, Berkeley, first edition: 1922.