

Article

# Quantum Type Indeterminacy in Dynamic Decision-Making: Self-Control through Identity Management

Ariane Lambert-Mogiliansky <sup>1,\*</sup> and Jerome Busemeyer <sup>2</sup>

<sup>1</sup> Paris School of Economics, 48 bd Jourdan, 75014 Paris, France

<sup>2</sup> Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47 468, USA;  
E-Mail: jbusemey@indiana.edu

\* Author to whom correspondence should be addressed; E-Mail: alambert@pse.ens.fr;  
Tel: +33-143136395.

Received: 29 February 2012; in revised form: 29 March 2012 / Accepted: 30 March 2012 /

Published: 15 May 2012

---

**Abstract:** The Type Indeterminacy model is a theoretical framework that uses some elements of quantum formalism to model the constructive preference perspective suggested by Kahneman and Tversky. In a dynamic decision context, type indeterminacy induces a game with multiple selves associated with a state transition process. We define a Markov perfect equilibrium among the selves with individual identity (preferences) as the state variable. The approach allows to characterize generic personality types and derive some **comparative** static results.

**Keywords:** indeterminacy; decision-making; self-control; identity

**JEL Classification:** D03, D81, C73

---

*“The idea of self-control is paradoxical unless it is assumed that the psyche contains more than one energy system, and that these energy systems have some degree of independence from each other.” ([1])*

## 1. Introduction

Everyone has probably seen an ambiguous picture. One famous example allows one to see either the profile of Sigmund Freud’s head or the naked body of a little woman. The remarkable thing is that one

cannot see both simultaneously. Both pictures are true, but they are incompatible. This recalls “Bohr complementarity” in quantum physics.<sup>1</sup> The complementarity principle states that some objects have multiple properties that appear to be contradictory. Sometimes it is possible to switch back and forth between different views of an object to observe these properties, but in principle, it is impossible to view both at the same time, despite their simultaneous existence.<sup>2</sup> Similarities between human sciences and quantum physics were early recognized by the founders of quantum mechanics, including Bohr and Heisenberg.<sup>3</sup> In particular, Bohr was heavily influenced by the psychology and philosophy of knowledge of Harald Høffding.<sup>4</sup> A fundamental similarity stems from the fact that in both fields, *the object of investigation cannot (always) be separated from the process of investigation*. Quantum mechanics and in particular its mathematical formalism were developed to respond to that epistemological challenge (see the Introduction in [5]). In our view this makes it fully legitimate to explore the value of the mathematical formalism of quantum mechanics in the study of human behavioral phenomena.

Over the last decade, scholars from social sciences, psychology, physics and mathematics have contributed to the development of a “quantum-like” decision theory based on the premises of (non-classical) indeterminacy (see e.g., [6–14]). This line of research has proved to be very fruitful in explaining a wide variety of behavioral phenomena ranging from cognitive dissonance to preference reversal, the inverse fallacy or the disjunction effect. A central feature of (non-classical) indeterminacy is, according to G. Mackey ([15]), that it places limitations on the system: “The laws of quantum mechanics place certain restrictions on the possible simultaneous probability distribution of various observables” (p. 61). Similarly, indeterminacy in decision theory captures cognitive limitations of the individual in the following sense. The individual is not simultaneously endowed with a preference order over all possible subsets of alternatives. Instead, as one elicits the individual’s preferences with respect to one subset of alternatives (in a choice experiment), his preferences with respect to another subset of alternatives (associated with an incompatible choice experiment) are modified so behavior can, for example, exhibit preference reversal.<sup>5</sup> This cognitive limitation implies that individual behavior is boundedly rational, in the sense that it is not consistent with the existence of a complete ordering over the universal set of alternatives.<sup>6</sup>

The starting point for our approach is that we depart from the classical dogma that individuals are endowed with preferences and attitudes that motivate their behavior. Instead, we propose that the motivational underpinning of behavior is intrinsically uncertain, *i.e.*, indeterminate. It is only at the

---

<sup>1</sup>Niels Bohr was one of the founders of quantum mechanics. After intense discussions with Heisenberg and Pauli, he introduced the fundamental concept of complementarity at the Côme conference in 1927, and this was followed by numerous publications.

<sup>2</sup>Ambiguous pictures do not feature all characteristics of Bohr complementarity. For a rigorous analysis of quantum-like phenomena in human perception see e.g., [2]. They study oscillations in bi-stable perception of the Necker cube.

<sup>3</sup>Heisenberg [3] distinguishes three regions of knowledge depending on the degree of separability between the object and the process of investigation. The second region corresponds to the case where we have non-separabilities. Heisenberg puts quantum physics in that region, together with psychology and biology.

<sup>4</sup>Bohr considered introspective psychology not as an illustration but as the paradigmatic description of the epistemological limitations of modern physics. [4]

<sup>5</sup>Preference ordering over different subsets may be complementary properties in the sense of Bohr complementarity.

<sup>6</sup>For a comparison between the behavior of a classical rational man and that of a type-indeterminate agent, see [14] Section 3.1.

moment the individual selects an action that a specific type (preference) is actualized. It is not merely revealed but rather determined in the sense that prior to the choice, there is an irreducible multiplicity of potential types. This idea is very much in line with Tversky and Simonson, according to whom: “There is a growing body of evidence that supports an alternative conception according to which preferences are often constructed—not merely revealed—in the elicitation process. These constructions are contingent on the framing of the problem, the method of elicitation, and the context of the choice” (in [16], pp. 525–526).

The basic model of static decision-making with Type-Indeterminate agents, the TI-model, is formulated in [14]. As we consider dynamic individual optimization, the TI-model induces a game among potential incarnations of the individual. In each period, these potential incarnations represent conflicting desires or propensities to act. We formulate the decision problem in terms of a game between a multiplicity of (one-period lived) players, the selves.<sup>7</sup> They are linked to each other through two channels: (i) the selves share a common interests in the utility of the future incarnations of the individual and (ii) they are connected to each other in a process of state transition (which captures indeterminacy). In each period, the current selves form intentions to act. One action is played by the individual but the whole profile of (intended) actions matters to tomorrow’s identity, because of the state transition process. This creates a strategic concern among contemporaneous selves. In particular, when the selves pool, the individual’s preferences are unchanged while if they choose different actions, preferences are modified. We define a Markov perfect equilibrium among the selves where the state variable is the individual’s identity. In our model, behavior affects future preferences (identity) and in particular a concern for identity (self-image) arises endogenously, because identity determines future expected utility. Choice behavior exhibits deviations from standard utility maximization. It is characterized by some degree of self-control: some selves may refrain from short-run gains (and pool with others) to secure a desirable identity. It can also feature dynamic inconsistency, because as preferences are modified, the choices made by the individual through time are not consistent with a stable preference order. The model delivers some of the predictions of Benabou and Tirole [18,19], in particular regarding the impact of the concern for identity on choice behavior. It also generates novel predictions. We characterize generic classes of personality/behavior: an unconflicted, weakly decisive but behaviorally stable character and a highly conflicted, strongly decisive and behaviorally unstable character.

We have recently witnessed renewed interest among prominent economic theorists in the issue of self-control and dynamic inconsistency in decision-making (see e.g., [20–24]). There is a significant theoretical literature pioneered by Strotz [25] dealing with various forms of dynamic inconsistency. A large part of this literature focuses on inconsistency that arises because the individual does not discount the future at a constant rate. Some form of myopia is assumed instead (e.g., quasi-hyperbolic discounting). Another approach to the planning problem, first proposed by Peleg and Yaari [26], models individual decision-making in terms of multiple selves. Various ways to model those selves and the interactions between them have been investigated. Fudenberg and Levine [23,24] develop a dual-self

---

<sup>7</sup>Although this paper uses elements of the quantum formalism in games, we are not dealing with so-called quantum games which study how the extension of classical moves to quantum ones can affect the analysis of the game. That approach consists in changing the strategy space, see for instance [17]. We model a game where the agent is characterized by (quantum) indeterminate preferences.

model of self-control with a long-term benevolent patient self and a multiplicity of impulsive short-term selves. This particular structure allows them to write the game as a decision problem and they can explain a number of behavioral paradoxes. In the present paper, we argue that the quantum approach to decision-making offers a suitable framework for McIntosh's paradox of self-control, because the indeterminacy of individual preferences is equivalent to the multiplicity of the selves (the potential eigentypes). Our approach contributes to the literature on self-control by investigating a mechanism of self-control based on identity management that uses (intrinsic) type indeterminacy. The technology for the evolution of identity reflects the dynamics of state transition under non-classical indeterminacy, and individual identity is the equilibrium outcome of the interaction between the selves. One contribution is that we formalize internal conflicts and explain features of self-management without using time preferences, which were the almost exclusive focus of earlier works on dynamic inconsistency. Moreover, we can connect to another branch of research related to identity and self-image, extensively investigated in psychology (especially in self-perception theory, see next section) and more recently in economics, see e.g., Benabou and Tirole [18,19].

The paper is organized as follows. In the next section we present some motivating puzzles and argue that our modeling is closely linked to the so-called self-perception theory in psychology. Next we present a general model of dynamic optimization by a type indeterminate agent. We illustrate the main features in an example. We define generic personality types in terms of the fundamentals of the model and derive some comparative static results. Finally, we discuss links between our model and the economic literature on self-control and identity.

## 2. Motivating Puzzles

The idea that an individual's choice of action (behavior) determines his inner characteristics (preference, attitudes and beliefs) rather than (exclusively) the other way around has been present in people's minds throughout history and has been addressed in philosophy, psychology and more recently in economics.

Nevertheless, the dominant view, particularly in economics, is based on a postulate: individuals are endowed with an identity (preferences, attitudes and beliefs) that explain their behavior. This postulate is hard to reconcile with a host of experimental evidence, including at the more basic level of perception. It has long been known that the perception of pain is only partly a function of the pain stimulus. Zimbardo *et al.* in [27] demonstrated that individuals who had volunteered to continue participating in an experiment using painful electric shocks, reported the shocks to be less painful and were physiologically less responsive than individuals who were given no choice about continuing. Valins and Ray [28] conducted an experiment where snake-phobic subjects were shown pictures of snakes and were falsely told that their heartbeat was calm. Subsequently, they exhibited significantly reduced fear of snakes. In another experiment, subjects were cued to identify the same physiological arousal as either anger or euphoria [29]. Cognitive dissonance experiments in e.g., the classic Carlsmith and Festinger experiment [30] also show how behavior affects attitudes. For a systematic review of experimental evidence (see [31]). All this evidence led Weick [32] to propose that: "Attribution and attitudes may follow behavior and not precede it". Similarly, Berkowitz [33] remarks: "We generally assume as a matter of course that the human being acts because of the wants arising from his understanding of the

environment. In some cases, the understanding may develop *after* stimuli have evoked the action, so that the understanding justifies but does not cause the behavior” (p. 308).

Psychologists have developed several theories to account for these experimental facts. According to the famous James-Lang theory of emotion, when an emotional event occurs, our behavioral reaction determines our subjective experience of the event (see [34]). Closely related is self-perception theory. As expressed by Bem [31], self-perception theory is based on two postulates: 1. “*individuals come to ‘know’ their own attitude and other internal states partially by inferring them from observations of their own behavior and/or the circumstances in which behavior occurs*; 2. *Thus the individual is functionally in the same position as an outside observer, an observer who must necessarily rely upon those same external cues to infer the individual inner state.*” (p. 2 in [31]). Self-perception theory does not clearly give up the classical postulate. Nevertheless, its own postulates are fully consistent with the hypothesis of (non-classical) indeterminacy which overturns the classical postulate of pre-existing identity, attitudes and preferences. With indeterminacy of the inner state, behavior (the action chosen in a decision situation, see below) shapes the state of preferences/attitudes through a state transition process (see next section). Indeterminacy means intrinsic uncertainty about individual identity so the individual may not know his *own* attitudes, preferences and beliefs. And as in self-perception theory, it is by observing his own action that he infers (learns) his state (of beliefs and preferences). While self-perception theory emphasizes the similitude between outside observation and self observation, quantum decision theory emphasizes the fact that observation is structured. As recognized in self-perception theory, inner states are not accessible without some training and instruments to measure them. To observe, one needs an “appropriate descriptor” ([31] p. 3). Such a “descriptor” includes “cues” that can be manipulated to obtain widely different perceptions (as in the experiment on anger versus euphoria cited above). This is consistent with the most basic feature of indeterminacy, namely that the property of a system does not pre-exist observation. Therefore different measuring instruments may give various incompatible but equally true accounts of the same state. As we shall see, this is also at the heart of the state transition process and delivers our theory of self-control.

### 3. The Model

We shall describe the dynamic decision problem as a simple separable dynamic game between the selves of an individual with the state variable identified as the identity (type) of the individual. The equilibrium concept we shall be using is that of Markov perfect equilibrium.

The kind of scenario we have in mind is a sequence of (at least) two consecutive decision situations (DS), as in the following example. Bob has just inherited some money from his aunt and the first decision is between buying state obligations or risky assets. The second decision is between a stay-at-home evening or taking his wife to a party. The two situations appeal to different but related (in a sense to be made clear below) type characteristics: the first *DS* calls upon his preference vis-à-vis risk: cautious ( $\theta_1$ ) or risk-loving ( $\theta_2$ ). The second situation calls upon his attitude towards others: ( $\tau_1$ ) egoistic or generous/empathetic ( $\tau_2$ ). An alternative scenario that brings us closer to Tirole and Benabou and to the literature on self-control, is to define DS1 as a choice between exercising or sleeping late. DS2 is a choice between watching a good movie or helping your mother with chores. The idea is that both *DS* involve a choice between tempting (immediate) gratification and more sophisticated satisfaction.

We next develop the general theory and illustrate it with the above mentioned example.

### 3.1. The Players

In each period, the individual faces a decision situation (*DS*)  $A^t$  corresponding to the finite set of available actions in period  $t$ . We restrict the one-period players' strategy set to pure actions. The possible preferences over the profiles of actions (one action for each self) are denoted by  $e_{M,i}$  where  $M$  defines the complete measurement corresponding to  $A^t$  (see below). Consider  $A^t = \{a_1, a_2\}$  and assume that there are (only) three possible preferences: prefer the action chosen by the other selves (and if the two others play different actions, randomize) or a strict preference for  $a_1$  ( $a_2$ ) irrespective of the others' choices.  $M \in \mathcal{M}$ , where  $\mathcal{M}$  is the set of all complete measurements, corresponds to an elicitation procedure that fully reveals the preferences in  $A_t$ . A choice in *DS*  $A^t$  is a coarse measurement e.g., in our example the choice of  $a_1$  by a self does not allow to distinguish between the pooling type and the  $a_1$  dominant action type if at least one of the other selves also chooses  $a_1$ . We refer to the  $e_{M,i}$  as the selves or the "eigentypes" of  $M$ .<sup>8</sup> They are the players in our game.

In each period  $t$  the individual is represented by his state or type (we use the terms interchangeably), a vector  $|s^t\rangle \in \mathbb{S}$ , where  $\mathbb{S}$  is a (finite)  $n$ -dimensional Hilbert space and the bracket  $|\cdot\rangle$  denotes a (ket) vector in Dirac's notation, which is standard when dealing with indeterminacy.<sup>9</sup> The eigentypes  $e_{M,i}$  of  $M$  are associated with the eigenvectors  $|e_{M,i}\rangle$  of the operator, which form a basis of the state space. The state vector can therefore be expressed as a superposition:<sup>10</sup>  $|s^t\rangle = \sum_{i=1}^n \lambda_i^t |e_{M,i}\rangle$ ,  $\lambda_i \in \mathbb{R}$ ,  $\sum_i (\lambda_i^t)^2 = 1$ , where  $e_{M,i}$  are the (potential) selves relevant to *DS*  $A_t$ . This formulation means that the individual cannot generally be identified with a single true self. He does not have a single true preference, instead he is intrinsically "conflicted", which is expressed by the multiplicity of potential selves.<sup>11</sup> The coefficient  $\lambda_i$ , also called amplitude of probability, provides a measure of the relative strength of potential self  $e_{M,i}$ , more precisely the square of the coefficient gives the probability that self  $e_{M,i}$  will determine the behavior of the individual in *DS*  $A_t$ . As a special case we have  $|s^t\rangle = \sum_i \lambda_i^t |e_{M,i}\rangle$  with  $\lambda_i = 0, i = 2, \dots, n$  implying  $|s^t\rangle = |e_{M,1}\rangle$ . In that special case, the individual is, at time  $t$ , identified with self  $e_{M,1}$ .<sup>12</sup> As we shall see below, if the individual preferences are fully determined in *DS*  $A^t$ , they are by necessity indeterminate in any non-compatible *DS*  $A^{t+1}$ .

We assume throughout the paper that there is common knowledge among the players (selves) about the current state, the utility function of all players and the state transition process (see below). As argued in the introduction, type indeterminacy implies bounded rationality *at the individual level* in the sense that individual behavior is not consistent with a well-defined complete ordering over the universal set of

<sup>8</sup>An eigentype corresponds to an eigenvalue of the operator.

<sup>9</sup>The mathematical concept of a Hilbert space generalizes the notion of Euclidean space. Its significance was underlined with the realization that it offers one of the best mathematical formulations of quantum mechanics. In short, the states of a quantum mechanical system are vectors in a certain Hilbert space, the observables are Hermitian operators on that space, and measurements are orthogonal projections

<sup>10</sup>A superposition is simply a linear combination such that the square of the coefficients add up to 1.

<sup>11</sup>In the human mind conflicting propensities to act co-exist until a choice is made. For instance a person may hesitate between two desserts: a chocolate cake or a frozen yogurt. With the actual choice of the frozen yogurt, she becomes a person capable of resisting the temptation to eat a chocolate cake. However, this identity can be reversed by later choices.

<sup>12</sup>In that special case, the individual's preferences in  $A_t$  pre-exist the measurement, *i.e.*, the act of choice.

alternatives. As we shall see in the next section, indeterminacy also means that preferences are unstable. However, we assume that the individual is aware of his own indeterminacy and acts consistently within the corresponding cognitive limitations. This hypothesis is captured in an assumption of rationality (in a way to be defined below) and common knowledge of rationality at the level of the selves. We do not claim that one should always expect such a degree of sophistication from the selves but our approach is a natural first step.<sup>13</sup>

### 3.2. Indeterminacy: Decision-Making as a State Transition Process

In each period, the selves form intentions to play and eventually *one* action is taken by the individual. Decision-making is modeled as the measurement of the preferences (see the revelation principle) and it is associated with a transition process from the initial state and (intended) actions to a new state. The rules governing the state transition process reflect the intrinsic indeterminacy of the individual’s type or preferences. It features the minimal perturbation principle defining a measurement operation which is formalized by the von Neumann projection postulate:<sup>14</sup> if the initial state is  $|s\rangle$  and the chosen action is  $a_1$  then the new state is the normalized projection of  $|s\rangle$  onto the eigenspace belonging to  $a_1$ .<sup>15</sup>

Formally, a transition process is a function from the initial state and (intended) actions to a new state. It can be decomposed into an outcome mapping  $\mu_A : \mathbb{S} \rightarrow \Delta A$  where  $\Delta A$  is the unit simplex of actions and a transition mapping  $\tau_{M,a} : \mathbb{S} \rightarrow \mathbb{S}$ . The first mapping defines the probability for the possible choices of action when an individual in state  $|s\rangle$  is confronted with *DS*  $A$ . The second mapping  $\tau_{M,a}$  indicates where the state transits as we confront the individual with *DS*  $A$  and obtain outcome  $a$ .

Let the initial state be  $|s^t\rangle = \sum_i \lambda_i^t |e_{M,i}\rangle$ . The standard Hilbert space formulation yields that if we, for instance, observe action  $a_1$ , the state transits onto:

$$|s^{t+1}\rangle = \sum_{j=1} \lambda'_j |e_{M,j}\rangle \tag{1}$$

where  $\lambda'_j = \frac{\lambda_j^t}{\sqrt{\sum_{k^t} (\lambda_k^t)^2 (s_k^* = a_j)}}$  and  $\sum_{k^t} (\lambda_k^t)^2 (s_k^* = a_j)$  is the sum over the probabilities for the selves who pool in choosing  $a_j$ . This is of course equivalent to Bayesian updating *i.e.*, the state transition seems purely informational. The value of this more general formulation comes when dealing with a sequence of non-commuting *DS*. To see that the formal equivalence breaks down, we have to express  $|s^{t+1}\rangle$  in terms of  $|e_{N,i}\rangle$  where  $N$  is the new (non-commuting) measurement in period  $t+1$  corresponding to *DS*  $A^{t+1}$  and  $|e_{N,i}\rangle$  are its eigenvectors. The eigenvectors of  $N$  also form an (alternative) basis of the state space. And this is where the above-mentioned correlations between selves from different periods enter into play. The correlations link the two sets of basis vectors: the eigenvectors of  $M$  can be written as linear combinations of the eigenvectors of  $N$  with the correlations as the coefficients of superposition—see below.

<sup>13</sup>As we shall see, the extent of the selves’ sophistication is modulated through the coefficient given to the utility of future incarnation.

<sup>14</sup>Or the postulate’s more stringent version defined by Luder.

<sup>15</sup>We talk about the “eigenspace” associated with an eigenvalue “a” of a measurement operator if the eigenvalue is degenerate *i.e.*, if several linearly independent vectors yield the same outcome of the measurement.

These correlations capture the extent of overlap between the states.<sup>16</sup> In a classical world *all* distinct *atomic* states are orthogonal. So in a classical world, either the type characteristics are mutually exclusive—so for instance Bob is either of the risk-loving type or of the cautious type—or the types can be combined—Bob can be of the risk-loving type *and* of the egoistic type. But in the latter case, the type characteristic “risk-loving” is not be a complete characterization *i.e.*, not an atomic state. The novelty with indeterminacy is that type characteristics can overlap in the sense that they are non-orthogonal atomic states. For instance, in our example the risk-loving type and the cautious type are orthogonal, but the risk-loving type and the egoistic type are not. Nevertheless, the three are complete descriptions of the individual *i.e.*, they are atomic states. The risk-loving type overlaps with the egoistic type. This means that if Bob is of the risk-loving type, there is some probability that in his second choice he will reveal egoistic preferences (with the complementary probability he reveals generous preferences) and his type will be modified, he will no longer be of the risk-loving type. Instead, he will be fully characterized as an egoistic type. And, if tested again with respect to cautious/risk-loving characteristics, the state will transit again and he may end up as a cautious type. The correlations are a measure of this overlap.

Let  $B_{MN}$  denote the basis transformation matrix that links the two non-compatible type characteristics  $M$  and  $N$ :  $|e_{M,i}\rangle = \sum_j \gamma_{ij} |e_{N,j}\rangle$  where  $\gamma_{ij}$  are the elements of the basis transformation matrix  $\gamma_{ij} = \langle e_{N,j} | e_{M,i} \rangle$ .<sup>17</sup> Substituting into Equation (1) and collecting the terms we write:

$$|s^{t+1}\rangle = \sum_j \left( \sum_i \lambda'_i \gamma_{ij} \right) |e_{N,i}\rangle = \sum_i \eta_i^{t+1} |e_{N,i}\rangle$$

According to Bohr’s rule the probability for eigentype  $|e_{N,1}\rangle$  (if the agent is confronted with  $DS A^{t+1}$  that (coarsely) measures type characteristics  $N$ ) is:

$$TP : p(e_{N,1} | |s^{t+1}\rangle) = \left( \sum_i \lambda'_i \gamma_{1i} \right)^2 \tag{2}$$

This is a crucial formula that captures the key distinction between the classical and the type indeterminacy approach. TP is *not* a conditional probability formula where the  $\gamma_{ij}^2$  are statistical correlations between the eigentypes at the two stages. The probabilities for the  $N$ -eigentypes depend on the  $M$ -eigentypes’ *play* in  $DS A^t$ . When no player chooses the same action, the choice of  $a_i^t$  separates out a single player (some  $e_{M,i}$ ), the sum in parentheses involves *one* term only. Whereas, when several players pool in choosing the same action, the term in parentheses involves several terms. As a consequence, the probabilities for the different players are given by the square of a sum, involving cross-terms called interference effects—and not the sum of squares (as we would have in a classical setting). Since the amplitudes of probability can be negative numbers, the interference effect may be negative or positive.

<sup>16</sup>In their remarkable book on quantum logic [35], Beltrametti and Cassinelli write: “In physics, the expression transition probability generally refers to dynamical instability. Our use of the term is not directly related to instability, rather, we follow von Neuman’s terminology. The transition probability between two states is meant to represent intuitively a measure of their overlapping. Actual transition from one state to another is triggered by a measurement.”

<sup>17</sup> $\langle e_{N,j} | e_{M,i} \rangle$  is a scalar product.

We note that the state transition process is deterministic by the above-mentioned von Neumann’s postulate, which says that under the impact of a measurement a pure state transits into another pure state. In this paper we are only dealing with pure types. If we observe  $a_1^t$  (as the result of applying  $A_t$ ) the state

$$|s^t\rangle = \sum_i \lambda_i^t |e_{M,i}\rangle \text{ transits onto } |s^{t+1}\rangle = \sum_{j=1} \lambda_j^t |e_{M,j}\rangle = \sum_i \eta_i^{t+1} |e_{N,i}\rangle$$

that is  $|s^{t+1}\rangle$  is a pure state. Yet predictions on the outcome of  $A^{t+1}$  are probabilistic because of indeterminacy, *i.e.*,  $|s^{t+1}\rangle$  is a superposed state.

### 3.3. Utility

When dealing with multiple selves, the question as to how to relate the utility of the selves (here the players) to that of the individual has no self-given answer.<sup>18</sup> We adopt the following definition of the utility of self (or player)  $e_{M,i}$  of playing of  $a_i^t$  when the  $-i$  other  $t$ -period players play  $\mathbf{a}_{-i}^t$

$$U_{e_{M,i}}(a_i^t) + \delta_{e_{M,i}} \sum_{i=t}^T EU(s^{i+1}(a_i^t, \mathbf{a}_{-i}^t; s^t | a^t = a_i^t)) \tag{3}$$

where  $a^t$  denotes the actual play of the individual.

The utility for  $e_{M,i}$  of playing  $a_i^t$  consists of two terms. The first is the utility in the current period evaluated by player  $e_{M,i}$ . This term only depends on the action chosen by  $e_{M,i}$ . The second is the expected utility of the individual evaluated by the future selves conditional on  $a^t = a_i^t$ . The second term depends indirectly on the whole profile of (intended) actions in the current period through the state transition process  $s^{t+1}(\mathbf{a}_i^t; s^t)$ . The next period expected utility is  $EU(s^{t+1}(\mathbf{a}^t; s^t)) = \sum_i \eta_i(a_i^t | \mathbf{a}^t) U(s^{i+1}(a_i^t; s^t))$ , where  $\eta_i(a_i^t | \mathbf{a}^t)$  is the coefficient of superposition relating to the next period *DS*. It is the weighted sum of the utility of all the possible resulting types following  $\mathbf{a}^t$ , where the weights are given after updating and expressing the new state according to TP. The possible resulting states  $s^{t+1}(a_i^t; s^t)$  are expressed in terms of the eigentypes relating to  $A^{t+1}$  and the expected utility of period  $t + 1$  is calculated given the optimal choice of eigentypes relating to period  $t + 1$ , *e.g.*,  $e_{N,j}$ .

The current action profile only influences tomorrow’s state, the summation term in Equation (3) can therefore be collapsed into a single term  $EU^T(s^{t+1}(\mathbf{a}^t; s^t)) = \sum_{i=t}^T EU^*(s^{i+1}(\mathbf{a}^t; s^t))$ . Which is the expected utility when all future selves in all periods play an equilibrium pure strategy.<sup>19</sup>

The utility can thus be written:

$$U_{e_{M,i}}(a_i^t; s^t) + \delta_{e_{M,i}} EU^T(s^{t+1}(\mathbf{a}^t; s^t | a^t = a_i^t)) \tag{4}$$

In each period the payoff relating to the history of play is captured by the state variable representing the current state or identity.

The utility function may recall a Bernoulli function in the following sense. With some probability, the self survives (his preferred action is played by the individual) and with the complementary probability,

<sup>18</sup>One reason is that while the selves are incarnations of the same individual, they are short-lived. Another is that they might not recognize the “legitimacy” of some future possible incarnations. For instance a current compassionate self may not value the utility of a future spiteful incarnation.

<sup>19</sup>For the case when there exist mutiple equilibria, we assume that the current selves share the same beliefs about which equilibrium is played.

he is “out of the game”. The formulation in Equation (4) means that he maximizes utility conditional on surviving. The probability for survival depends on the initial coefficients of superposition and his own and other selves’ choices. But the selves do not take that into account. The approach is justified on the following grounds: being “out of the game” cannot be valued. The self ceases to “exist”, which is neither good nor bad. In other words, there is no reason to assume that selves have a “survival instinct”; they are simply mental constructs. A self is defined as rational when he maximizes his conditional utility, which is well-defined for any sequence of  $DS$ .

### 3.4. The Equilibrium

In each period, the current selves move simultaneously. They know the current state resulting from the previous (actual and intended) play. We have common knowledge among the selves about the payoff functions of all current and future selves and common knowledge of rationality. The selves’ payoffs are functions of the current actions and the current state as defined in the previous section.<sup>20</sup> Together, this means that we are dealing with a separable dynamic game of complete information and that it seems most appropriate to restrict ourselves to Markov strategies: a strategy for a self is a function  $\mathbb{S} \rightarrow A^t$  from the current state to the set of actions available at period  $t$ . We shall accordingly focus on Markov perfect equilibria.

**Definition 1** A Markov perfect equilibrium of the game is characterized by  $a_i^{t*}$  :

$$a_i^{t*} = \arg \max_{a_i \in A^t} U_{e_{M,i}}(a_i^t; s^t) + \delta_{e_{M,i}} \sum_{\tau=t+1}^T EU^*(s^\tau(\mathbf{a}^t; s^t))$$

in all periods  $t = 1, \dots, T$  and for all  $e_{M,i}$ ,  $M \in \mathcal{M}$ ,  $i = 1, \dots, n$ .

So we see that a self “only” needs to worry about his current utility and the expected utility value of his action via the resulting type. The equilibrium is found by backward induction in a standard way.<sup>21</sup> restrict ourselves to pure strategies. The novelty lies in the technology for the state transition process which captures indeterminacy. So, in particular, the state variables are the preferences themselves and they evolve in a non-monotonic way, reflecting the dynamics of measurement operations and the correlations between non-commuting  $DS$ .

**Remark 1** For the case where all  $DS$  commute with each other, the model is of an individual who does not initially know his preferences and learns through Bayesian updating as he observes the actions he takes.

If all  $DS$  commute, the state variable evolves through Bayesian updating. The individual eventually learns who he is and behaves as a classical decision-maker who maximizes discounted expected utility. In the TI-model, the concern for identity arises exclusively as a consequence of the non-commutativity of successive  $DS$ . The general case is one where some  $DS$  commute and some do not. Below, we focus

<sup>20</sup>Maskin and Tirole [36] develop a general approach to Markov perfect equilibrium where strategies may depend on history in a more elaborate way. A distinguishing feature of a quantum state is that a measurement erases information about the previous state. In other words, all relevant information for predicting the outcome of any measurement is contained in the current state. The history of the state has no relevance. This feature invites the restriction to simple Markov strategies.

<sup>21</sup>Although we know that a MPE exists in mixed strategies (see theorem 13.1 in [37]), we have no proof of existence for the present case, where we

on non-commuting *DS*, which allows us to address the issue of identity management in each period. But we should keep in mind that the kind of preference instability described in the next section does not apply within a sequence of commuting *DS*.

**Definition 2** *We say that the MPE is characterized by self-control when contemporaneous selves with conflicting short-run preferences pool to select the same action.*

When the action set is sufficiently rich to fully sort out the preferences, all selves have “conflicting” preferences with respect to the short-run choice. If the MPE is characterized by pooling, some selves must be exercising self-restraint, refraining from immediate reward for the sake of the individual future utility—this is an instance of self-control. In a standard *DS*, the set of actions is limited relative to the possible preferences (because a *DS* is generally a coarse measurement); we talk about self-control only when selves with short-run conflicting interests with respect to the *DS* pool. In the next section, we consider an example where the set of actions is sufficiently rich relative to preferences.

**Remark 2** *For the special case with  $\delta_{e_{M,i}} = 0$  for all selves in all periods, we are back in the basic TI-model. There is no self-control. For  $\delta_{e_{M,i}} \neq 0$  for some selves in some periods, the equilibrium path of action may exhibit some degree of self-control. The model suggests a classification of individual traits and behavior, as we show below.*

For the case where the selves are short-sighted (or unaware of the impact of action on the future *i.e.*, the case of unsophisticated selves), we are back in the simple decision-making model formulated in [14]. This has been used to explain behavioral anomalies in decision theory from cognitive dissonance to framing effects and preference reversal.

The case with  $\delta_{e_{M,i}} = 1$ , for all selves in all periods, is interesting because a classical agent would not face any self-control problem. In contrast, for a type-indeterminate decision-maker, the issue of self management arises because today’s intended actions affect future identity - see the example below. For the case where all but one self in each period are short-sighted ( $\delta_{e_{M,i}} = 0$ ), we have a model with the dual structure recalling Fudenberg and Levine, but the means of controlling the behavior are very different (see Discussion section below). We leave to future research the investigation of a type-indeterminate individual with such a dual structure.

#### 4. An Illustrative Example

Bob faces two consecutive non-commuting decisions: *DS1* with action set  $\{a_1, a_2\}$  and *DS2*:  $\{x_1, x_2\}$ . The story is as follows. Bob just inherited some money from his aunt. The first decision situation involves a choice between buying state obligations ( $a_1$ ) and risky assets ( $a_2$ ). The second *DS* is a choice between a stay-at-home evening ( $x_1$ ) and taking his wife to a party ( $x_2$ ). The type characteristic related to *DS1* has two values (eigentypes): cautious ( $\theta_1$ ) and risk loving ( $\theta_2$ ). In *DS2*, the type characteristic has two values as well: ( $\tau_1$ ) egoistic *versus* generous/empathetic ( $\tau_2$ ). We have deliberately selected two *DS* that are independent from each other in the sense that there is no complementarity or substitutability between the choices in *DS1* and *DS2*.

Below, we define the utility associated with the different choices. We assume that  $\tau_2$  experiences a high utility from  $x_2$  while  $\tau_1$  experiences a low utility whatever he chooses. In other words, it is better for Bob to be of the  $\tau_2$  type. We now provide the classical representation of the decision problem.

4.1. Classical Optimization

We first define the set of types under the classical assumption that all type characteristics are compatible. Since each type characteristic has two values, Bob may be any of the following four types  $\{\theta_1\tau_1, \theta_1\tau_2, \theta_2\tau_1, \theta_2\tau_2\}$ . The utility associated with the different options is given by Tables 1 and 2.

**Table 1.** First period payoffs.

$a_1$	$a_2$
$U(a_1; \theta_1\tau_1) = U(a_1; \theta_1\tau_2) = 4$	$U(a_2; \theta_1\tau_1) = U(a_2; \theta_1\tau_2) = 2$
$U(a_1; \theta_2\tau_1) = U(a_1; \theta_2\tau_2) = 2$	$U(a_2; \theta_2\tau_1) = U(a_2; \theta_2\tau_2) = 3$

Only the  $\theta$  value matters for the  $a$ -choice.

**Table 2.** Second period payoffs.

$x_1$	$x_2$
$U(x_1; \theta_1\tau_1) = U(x_1; \theta_2\tau_1) = 2$	$U(x_2; \theta_1\tau_1) = U(x_2; \theta_2\tau_1) = 0$
$U(x_1; \theta_1\tau_2) = U(x_1; \theta_2\tau_2) = 1$	$U(x_2; \theta_1\tau_2) = U(x_2; \theta_2\tau_2) = 8$

Here only the  $\tau$  value matters for the  $x$ -choice.

The tables above give us immediately the optimal choices:

$$\begin{aligned} \theta_1\tau_1 &\rightarrow (a_1, x_1) & \theta_2\tau_1 &\rightarrow (a_2, x_1) \\ \theta_1\tau_2 &\rightarrow (a_1, x_2) & \theta_2\tau_2 &\rightarrow (a_2, x_2) \end{aligned}$$

Using the values in Tables 1 and 2, we note that type  $\theta_1\tau_2$  achieves the highest total utility of 12. The lowest utility is achieved by  $\theta_2\tau_1$ .<sup>22</sup> In the classical setting, Bob knows his type but an outside observer only has some prior distribution over the four possible types, so Bob’s choice is uncertain from that perspective. We note that as expected, since the two  $DS$  are independent, the dynamic choice problem boils down to a sequence of two static decision problems.

4.2. A TI-Model of Dynamic Optimization

By definition, subjecting Bob to the  $a$ - choice is a measurement of his  $\theta$  characteristics,  $\theta \in: \{\theta_1, \theta_2\}$ . The type characteristic relevant to  $DS2$  is  $\tau, \tau \in \{\tau_1, \tau_2\}$ . We assume that the two  $DS$  are complete measurements so the type space under consideration is two-dimensional. Since the two  $DS$  do not commute, we can write:

$$\begin{aligned} |\theta_1\rangle &= \alpha_1 |\tau_1\rangle + \alpha_2 |\tau_2\rangle \\ |\theta_2\rangle &= \beta_1 |\tau_1\rangle + \beta_2 |\tau_2\rangle \end{aligned}$$

<sup>22</sup>We assume here that we can compare the utility of the different types of Bob. This goes beyond the standard assumptions in economics that preclude interpersonal utility comparisons. But it is in line with interpersonal comparisons made in the context of social choice theory.

where  $\begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{pmatrix}$  is a rotation matrix  $\alpha_1^2 + \alpha_2^2 = 1 = \beta_1^2 + \beta_2^2$ . We let  $\alpha_1 = \beta_2 = \sqrt{0.3}$  and  $\alpha_2 = -\beta_1 = \sqrt{0.7}$ .<sup>23</sup>

Since the two type characteristics are incompatible (*DS1* and *DS2* are non-commuting operators), the set of players is  $N : \{\theta_1, \theta_2, \tau_1, \tau_2\}$ . The  $\theta_i$ -players have action set  $\{a_1, a_2\}$  and they play at stage 1. At stage 2, it is the  $\tau_i$  players' turn, and they have action set  $\{x_1, x_2\}$ . There is an initial state  $|s^0\rangle = \lambda_1 |\theta_1\rangle + \lambda_2 |\theta_2\rangle$ ,  $\lambda_1^2 + \lambda_2^2 = 1$  with  $\lambda_1 = \lambda_2 = \sqrt{.5}$ . The transition process is defined by the rule for updating and the correlations between players at different stages  $(\alpha_1, \alpha_2, \beta_1, \beta_2)$  see section 3.2 above. If, for instance,  $a_1$  is the optimal choice for  $\theta_1$  and  $a_2$  the optimal choice for  $\theta_2$ , we know that if  $a_1$  is played, the state transits from  $|s^0\rangle$  to  $|s^1\rangle = |\theta_1\rangle = \alpha_1 |\tau_1\rangle + \alpha_2 |\tau_2\rangle$ .

The utility of the  $\tau$  players is as follows:

$$\begin{aligned} U_{\tau_1}(x_1) &= 2 & U_{\tau_1}(x_2) &= 0 \\ U_{\tau_2}(x_1) &= 1 & U_{\tau_2}(x_2) &= 8 \end{aligned}$$

and according to Equation (4) the utility of the  $\theta$  players is:

$$EU(a_{\theta_i}; \theta_i, s^0) = U_{\theta_i}(a_{\theta_i}) + \delta_{\theta_i} EU^*(s^1(a_{\theta_i}, a_{\theta_2}; s^0))$$

where we set  $\delta_{\theta_1} = \delta_{\theta_2} = 1$  for simplicity. The numerical values of the first term are given:

$$\begin{aligned} U_{\theta_1}(a_1) &= 4 & U_{\theta_1}(a_2) &= 2 \\ U_{\theta_2}(a_1) &= 2 & U_{\theta_2}(a_2) &= 3 \end{aligned}$$

To solve for the equilibrium, we proceed by backward induction to note that since the “world ends after *DS2*”,  $\tau_1$  chooses  $x_1$  and  $\tau_2$  chooses  $x_2$  (as in the classical model). Next, we fix the strategy of eigentype  $\theta_1$ ; let us say he chooses  $a_1$ .<sup>24</sup> What is  $\theta_2$ 's optimal choice? If he chooses  $a_2$ , the resulting state (relevant to his conditional utility) after *DS1* is  $|s^1\rangle = |\theta_2\rangle$ . The utility, in the first period, associated with the choice of  $a_2$  is  $U_{\theta_2}(a_2) = 3$ . In the second period, Bob's type is  $|\theta_2\rangle = \beta_1 |\tau_1\rangle + \beta_2 |\tau_2\rangle$  which, given what we know about the optimal choice of  $\tau_1$  and  $\tau_2$ , yields an expected utility of  $\beta_1^2 [U(x_1; \tau_1) = 1] + \beta_2^2 [U(x_2; \tau_2) = 8] = .7 + 8(.3) = 3.1$ . The total (for both periods) expected utility from playing “ $a_2$ ” for  $\theta_2$  is

$$EU(a_2; \theta_2, s^0) = 3 + 3.1 = 6.1$$

This should be compared with the utility, for  $\theta_2$ , of playing “ $a_1$ ”, in which case he pools with  $\theta_1$ ; the resulting type in the first period is the same as the initial type *i.e.*,  $|s^1\rangle = |s^0\rangle = \lambda_1 |\theta_1\rangle + \lambda_2 |\theta_2\rangle$ . The utility of playing  $a_1$  is  $U_{\theta_2}(a_1) = 2$  in the first period plus the expected utility of the second period. To calculate the latter, we first express the type vector  $|s^0\rangle$  in terms of  $|\tau_i\rangle$  eigenvectors:

$$|s^0\rangle = \lambda_1 (\alpha_1 |\tau_1\rangle + \alpha_2 |\tau_2\rangle) + \lambda_2 (\beta_1 |\tau_1\rangle + \beta_2 |\tau_2\rangle) = (\lambda_1 \alpha_1 + \lambda_2 \beta_1) |\tau_1\rangle + (\lambda_1 \alpha_2 + \lambda_2 \beta_2) |\tau_2\rangle$$

<sup>23</sup>Remember that the coefficients of superposition are amplitudes of probability which can take negative values, and that Bohr's rule calls for squaring them to obtain the probability for the corresponding eigentype.

<sup>24</sup>Note that the assumption of “ $a_1$ ” is not fully arbitrary since  $a_1$  gives a higher utility to  $\theta_1$  than  $a_2$ . However, we could just as well have investigated the best reply of  $\theta_1$  when assuming that  $\theta_2$  chooses  $a_2$ . See below and Note 22 for a justification of our choice.

The expected utility for the second period is calculated taking the optimal choice of  $\tau_1$  and  $\tau_2$  and the probabilities for  $\tau_1$  respectively  $\tau_2$ , *i.e.*, the square of the coefficients in the formula above:

$$(\lambda_1^2\alpha_1^2 + \lambda_2^2\beta_1^2 + 2\lambda_1\alpha_1\lambda_2\beta_1) 1 + (\lambda_1^2\alpha_2^2 + \lambda_2^2\beta_2^2 + 2\lambda_1\alpha_2\lambda_2\beta_2) 8 = 0.042 + 7.664 = 7,706$$

which yields:

$$EU(a_1; \theta_2, s^0) = 2 + 7,706 = 9,706 > EU(a_2; \theta_2, s^0) = 3 + 3.1 = 6.1$$

So we see that because of type indeterminacy and the associated state transition process, it is optimal for pure type  $\theta_2$  to forego a unit of utility in *DS1* and play  $a_1$  (instead of  $a_2$  as in the classical model). There is a gain for  $\theta_2$  in preserving the superposition because the interference effect ( $2\lambda_1\alpha_2\lambda_2\beta_2 \approx 0.46$ ) increases the probability of  $\tau_2$ . It can also be verified that given the play of  $\theta_2$  it is indeed optimal for  $\theta_1$  to choose  $a_1$ . Hence, pooling on  $a_1$  is part of a Markov perfect equilibrium.<sup>25</sup>

The multiple-self game is illustrated in Figure 1, where only the selves' choice in *DS1* and Nature's associated move are explicit.<sup>26</sup> The MPE is represented by a dotted line. It exhibits self-management. The interpretation is that Bob's  $\theta_2$  type understands that buying risky assets appeals to his risk-loving self, which makes him tense. He knows that when he is tense, his egoistic self tends to take over. So in the evening he is very unlikely to feel the desire to please his wife - his thoughts are simply elsewhere. But Bob also knows that when he is in the empathetic mood he always experiences deep happiness when pleasing his wife. So his risk-loving self may be willing to forego the thrill of doing risky business in order to increase the chance of achieving higher overall utility.

**Figure 1.** Multiple Selves game tree.



4.3. *Generic Classes of Behavior*

The two-types, two-actions and two-periods case allows us to illustrate some basic comparative statics results. The TI-model invites us to distinguish between two situations characterized by the sign of the interference effect applying to the high utility option. Interference effects are the signature of indeterminacy. When the individual is in the superposed state  $|s^0\rangle$ , both  $\theta$  eigentypes are simultaneously present in his mind and they interact. The interference effect captures the impact of the interaction between the  $\theta$  eigentypes in the determination of the probabilities for the outcome in *DS2*. The propensity to be of the type  $\tau_2$ , who experiences the high second-period utility, is present in both  $\theta_1$  ( $\alpha_2|\tau_2$ ) and  $\theta_2$  ( $\beta_2|\tau_2$ ). We have positive interference effects when those propensities reinforce each other and increase the chance that the superposed individual will turn out to be of that type. The sign of the interference effect depends on the operators associated with the decisions. These operators and the correlations between them are structural properties of the state space. Our view is that these properties

<sup>25</sup>The equilibrium may not be unique. The inner game is a coordination game. Since the preferences of  $\theta_2$  are more in line with those of  $\tau_2$  (high correlation), it seems that a natural selection would lead to coordination on  $a_2$  rather than  $a_1$ .

<sup>26</sup>When the selves pool, Nature's associated move is a null measurement (Nulmst). It reveals no information about the state, which is left unaffected.

capture neurological and psychological regularities common to all individuals, or at least a large group of them (e.g., from the same sociocultural background). Whether interference effects are positive or negative is an empirical question.<sup>27</sup> While these features are common to all individuals, each individual is characterized by his state, a vector in a potentially very high dimensional space. Moreover, individuals can differ in the intensity of the utility experienced by their selves and in the value their selves put on the utility of future incarnations. We next investigate how these characteristics, together with type indeterminacy, give rise to patterns of behavior or “personality traits”.

#### 4.3.1. Self-Control by Inner Agreement

In this section we assume, as in the example, that the interference effect (IE) favors the high utility option  $x_2$  (and automatically reduces the probability of the low utility alternative). As we shall see, a positive IE favors behavioral (and intertemporal) consistency. Assume that we have:

$$EU(\theta_2) < EU(\theta_1) < EU(s^0) \quad (5)$$

where  $EU(s)$  is the expected utility in period 2 when the state is  $s$ . Remember that the first  $DS$  is a complete measurement of type characteristics  $\theta$ <sup>28</sup> so we must have:

$$U_{\theta_1}(a_1) > U_{\theta_1}(a_2) \text{ and } U_{\theta_2}(a_1) < U_{\theta_2}(a_2)$$

*i.e.*, the  $\theta$ -selves have conflicting short-run interests.

When considering a sequence of two non-commuting  $DS$ , the model distinguishes between two classes of individuals: the balanced individual, an individual whose selves manage to agree on a common choice, and the conflicted individual, whose selves make separating choices, more precisely:

**Definition 3** *A balanced individual is characterized by an MPE that is a pooling equilibrium. It obtains whenever:*

$$U_{\theta_1}(a_1) + \delta_1 EU(\theta_1) \leq U_{\theta_1}(a_2) + \delta_1 EU(s^0) \quad (6)$$

*and/or:*

$$U_{\theta_2}(a_2) + \delta_2 EU(\theta_2) \leq U_{\theta_2}(a_1) + \delta_2 EU(s^0) \quad (7)$$

*Otherwise, the individual is conflicted i.e., his inner equilibrium is characterized by separation.*

Equations (6) and (7) capture the selves' incentives to refrain from choosing their preferred action (exerting self-restraint) when the other self chooses his preferred first-period action. When an inequality is falsified it is a dominating strategy for that self to choose his preferred first-period action. Since we have a conflict of interest, when neither of them holds, the choices are separating. This means that if invited to choose, the individual promptly incarnates either one or the other self. He shows clear-cut preferences, determination. This also means that the first-period action triggers state transition onto one of the eigentypes *i.e.*, identity is modified. As a consequence, behavior will exhibit inconsistency (e.g., preference reversal). So this suggests that individuals who are quite extreme in their judgment and have clear-cut preferences also exhibit behavioral inconsistency. The identity of a conflicted individual

<sup>27</sup>This is also the case in quantum mechanics.

<sup>28</sup>This means that when considered in isolation,  $DS1$  separates between the  $\theta$ -types.

jumps from one period to another with the decision made. Remember that this applies to non-commuting  $DS$ . So in a sequence  $DSI-DS3-DSI$  where  $DS3$  commutes with  $DSI$ , an individual that we characterize as conflicted will not exhibit any behavioral instability or inconsistencies.

The balanced individual is characterized by selves who are willing to reach an agreement; they make a pooling choice. This occurs at the expense of one of the selves who chooses to forego his preferred option in period 1. This is an instance of self-control. The balanced individual has no clear-cut preferences. We could say that he retains the freedom to value options from different perspectives. The pooling equilibrium obtains when the inequality in (6) and/or in (7) are true. We have pooling on  $a_1$  for the case, at least, where Equation (7) holds. When both inequalities hold, we could have pooling on either action. But since  $\theta_1$  is closer (highly correlated) to  $\tau_2$  and there is agreement on the advantage of identity  $\tau_2$ , it is reasonable to expect pooling on  $a_1$ . For the case where only Equation (6) holds, the MPE yields pooling on  $a_2$ . Interestingly, in two of the three cases, the pooling MPE yields the “good” action  $a_1$  in period 1. But this is not always the case. It may be that the “good” or more forward-looking self chooses to refrain from his preferred action—when Equation (6) does not hold. This may capture a situation where the individual feels that being too demanding with himself might backfire. In the example, if the cautious type insists on being cautious, there is a 50% chance that he becomes a risk-loving type who will have a high chance of being egoistic, which is costly since he will then only get a low second-period utility.

A pooling MPE triggers no state transition. If the selves pooled in all periods, the individual would simply behave as a an individual endowed with stable but stochastic preferences. He would not qualify as behaviorally or dynamically inconsistent.

We would like to emphasize that our model features self-control by *means of identity management*. In this respect, we stand closer to Benabou and Tirole [19]. In particular, we do not address the question related to taking actions (commitment) to limit future behavior as in Gul and Pesendorfer [20] and Fudenberg and Levine [23]. In the next section we briefly return to this question.

Definition 3 allows us to derive some simple comparative statics. For that purpose we write the inequalities in Definition 3 as follows:  $U_{\theta_i}(a_i) - U_{\theta_i}(a_j) \leq \delta_i [EU(s^0) - EU(\theta_i)]$ .

**Proposition 1**

1. The larger  $\delta_i$ ,  $i = 1, 2$  the more likely it is that we are dealing with a balanced type.
2. The larger the interference effect, the larger  $EU(s^0) - EU(\theta_i)$  and the more likely it is that we have pooling on the other self's preferred action.
3. The larger  $[U_{\theta_i}(a_i) - U_{\theta_i}(a_j)]$ , the more likely it is that the individual will turn out to be a conflicted person.

The proof follows from Definition 3 and is commented below.

1. The coefficient  $\delta_i$  captures the weight put by self  $\theta_i$  on the individual's identity and associated utility relative to the self's utility in the current period. Quite naturally, it also reads as a discounting factor. An individual composed of impatient selves (so both Equations (6) and (7) are violated) behaves erratically as a conflicted person. The more patient and/or concerned by identity the selves are, the more likely it is that we are dealing with a balanced individual. This is consistent with the comparative statics of Benabou and Tirole ([19], Proposition 2.c) on the effect of the salience of identity.
2. When the interference effect is large, this means that period 1's action can have a large impact on identity. The interference effect is preserved when the types pool but it is lost when they separate. In our context, the IE increases the probability of the type associated with the high utility alternative; this creates incentives to pool. Note further that we have  $EU(s^0) - EU(\theta_2) > EU(s^0) - EU(\theta_1)$ , because the separating action  $a_2$  induces identity  $\theta_2$  associated with low probability for  $\tau_2$ , while action  $a_1$  induces state  $\theta_1$ , which is highly correlated with  $\tau_2$ . This means that when preserving indeterminacy is optimal, the individual behaves in a way consistent with intuition. So, for instance, if the cost of foregoing the preferred alternative is the same for both selves, an individual who exerts self-control through identity management will choose in the first period the action preferred by the type highly correlated with the desired second-period identity. This is consistent with the kind of reinforcement effect found in Benabou and Tirole ([19], Proposition 2.d).
3. The utility difference  $[U_{\theta_i}(a_i) - U_{\theta_i}(a_j)]$  is the cost for self  $\theta_i$  of foregoing the preferred action. The magnitude of that difference captures the degree of conflict between the selves. Not surprisingly, when the potential selves have strong preferences, they will have a hard time agreeing (pooling) and the individual is likely to behave as a conflicted individual.

Positive IE with respect to the type associated with the high utility alternative means that indeterminacy with respect to *DSI* is generally (for all first-period selves) advantageous for the individual.<sup>29</sup> It is not surprising therefore that the MPE exhibits pooling, *i.e.*, preserves indeterminacy in many cases. Under these circumstances, we are likely to observe a significant extent of individuals behaving as balanced persons. We also observed that pooling occurs most of the time on the action preferred by the first-period eigentype most closely related to the preferred eigentype in period 2. This

<sup>29</sup>With positive IE we may also have:

$$EU(\theta_2) < EU(s^0) < EU(\theta_1)$$

if, e.g., the initial state is strongly skewed toward  $\theta_2$  e.g.,  $\lambda_2 = \sqrt{0.9}$ . The only pooling equilibrium would be on  $a_1$  and the only condition would be that Equation (6) is not verified. This shows that the magnitude of the IE depends on the initial state.

implies that there is high probability that the individual behaves as a cautious and altruistic person. But because of indeterminacy, the possibility that the individual “lapses” into poor behavior never disappears.

The correlation between  $DS$  and the associated interference effect can be viewed as the hardware of the mind. As mentioned earlier, just as in quantum mechanics we have no rigorous theory to derive those correlations. One conjecture is that preference traits that are in some sense “close” yet distinct are expected to be associated with non-commuting  $DS$ . So, for instance, cautiousness and altruism are neither identical nor orthogonal; they could instead be described as “overlapping”. As far as the sign of IE is concerned, we could expect the hardware of the human mind to evolve toward an efficient structure from a survival point of view.<sup>30</sup>

When considering the mind as a whole, we are likely to have some type characteristics associated with operators characterized by positive IE and others with negative IE. So it is relevant to ask: what behavior do we expect when the relevant interference effect is negative for the high utility alternative? In the next sub-section we briefly consider such a case.

### Negative Interference Effect: Agreeing to Disagree

Consider the case when:

$$EU(t) < EU(\theta_2) < EU(\theta_1) \quad (8)$$

which obtains in the example by inverting the signs of  $\beta_1$  and  $\alpha_2$ . The selves’ incentives are described unambiguously by the inequalities:

$$U_{\theta_1}(a_1) + \delta_1 EU(\theta_1) > U_{\theta_1}(a_2) + \delta_1 EU(t) \quad (9)$$

and:

$$U_{\theta_2}(a_2) + \delta_2 EU(\theta_2) > U_{\theta_2}(a_1) + \delta_2 EU(t) \quad (10)$$

This implies that both types prefer separation in  $DSI$ . We have a case of “agreeing to disagree”. Identity management concerns do not alter the selves’ short-run incentives. In the terms defined in the previous section, we say that with respect to  $DS$  linked by negative interference effects, the individual behaves as a conflicted person whose identity keeps on changing, so he exhibits intertemporal and behavioral inconsistencies.

We thus find that identity management concerns cannot promote self-restraint and consistency when IE are negative. In such cases, we may like to consider actions that limit future behavior. But that is outside the scope of the present paper.

## 5. Discussion

In this section we discuss the relation between the present work and some of the literature in economics. As mentioned in the Introduction, there exists a vast theoretical literature pioneered by Strotz [25] dealing with various types of time inconsistency. A large part of this literature has focused on inconsistency deriving from the fact that the individual does not discount the future at a constant

<sup>30</sup>In other words, we might expect the correlations that characterize the human mind to be the object of evolution (as other features of the human being are). This would mean that it tends to select correlation patterns that give a better chance of adapting to the environment.

rate. One of this paper's contributions is to demonstrate that there are other sources of inner conflict *i.e.*, not related to time preferences. A type-indeterminate individual is in each period characterized by a multiplicity of conflicting selves (competing desires). All selves are equally rational and care about the future expected utility of the individual. We formalize the "inner bargaining" formulated by Ainslie [38] as a sequential game and characterize the circumstances when individual behavior exhibits preference instability and intertemporal inconsistency. Our approach also differs from the economic literature on self control in another important respect. We do not consider actions whose primary objective is self-control, such as putting one's money into an account that is costly to access in order to limit one's spending opportunities. An important contribution of e.g., Gul and Pesendorfer [20] and Fudenberg and Levine [23] is to provide a rationale for such behavior. The present paper does not address this issue. One reason is that we may expect a commitment decision to commute with decisions related to immediate responses to temptations. Those decisions appeal to separate functions in the brain *i.e.*, higher cognitive functions as opposed to more visceral emotions and desires. If that is the case, which can be established in experiments, some of the reasoning of e.g., Fudenberg and Levine [23] would simply carry over. From the perspective of indeterminacy, a central question would instead be related to agenda setting. Indeed, the choice of whether or not to confront a *DS* about commitment is an important one which has consequences for identity.<sup>31</sup>

The type indeterminacy approach brings us close to the work by Benabou and Tirole ([18,19,39]), who write: "When contemplating choices, they then take into account what kind of a person each alternative would make them and the desirability of those self-views" ([19], p. 806–807). They continue: "Two related forms of behavioral instability are history dependence and non-monotonicity. When a person has been induced to behave prosocially or selfishly, or just provided with signals presumed to be informative about his morality, *his choices in subsequent, unrelated interactions are significantly affected*. Moreover, this reaction *sometimes amplifies* the original manipulation, and is *sometimes in opposition* to it." ([19], p. 810). A feature common to Benabou and Tirole's approach and the TI-model is that today's behavior affects tomorrow's identity (or self-image) *i.e.*, effective preferences. With type indeterminacy, individual identity is subject to a state transition process, so that future identity is a function of past actions. In Benabou and Tirole, the basic mechanism is incomplete information about own preferences associated with incomplete recall and incomplete self-control. More precisely, they depart from homo economicus by assuming instead (1) imperfect self-knowledge; (2) imperfect recall; (3) imperfect willpower. With these three imperfections they can derive the value of self-esteem (concern for identity) and self-monitoring behavior and accommodate intertemporal inconsistency in behavior.

As with the postulates of self-perception theory (see Section 2), we argue that the three assumptions in Benabou and Tirole are in many respects equivalent to giving up the classical dogma of a pre-existing (deterministic) individual identity and replacing it by indeterminacy. Indeterminacy implies imperfect knowledge because of intrinsic uncertainty: there is no set of "true preferences" (to be learned). Instead, an individual is represented by a superposition of potential types. Indeterminacy implies imperfect recall because no type is the true type forever. The (preference) state keeps changing with the action taken, so yesterday's correctly-inferred information about oneself may simply not be valid tomorrow. Indeterminacy implies "imperfect willpower" because it involves selves that are multiple

<sup>31</sup>It also has implications for other decisions that do not commute with the commitment decision.

both simultaneously (multiplicity of potentials) and dynamically (by force of the non-commutativity of decision situations). Therefore, there are necessarily conflicting desires and issues of self-control and self-monitoring. Moreover, in a world of indeterminate agents, actions aimed at shaping one's identity are fully justified from an instrumental point of view (it determines future expected utility). In particular there is no need to add any additional concerns for self-image (as in Benabou and Tirole), or diagnostic utility (as in [40]). The TI-model provides a simple and rigorous setting relying on one single departure from the standard setting.<sup>32</sup> Some of our comparative statics results (see Section 4.3) are similar to those in Benabou and Tirole and consistent with a host of empirical data, including those mentioned in Section 2. Our contribution is to propose an alternative explanation in terms of a fundamental characteristic of the mind: its intrinsic indeterminacy.<sup>33</sup>

## References

1. McIntosh, D. *Foundation of Human Society*; University of Chicago Press: Chicago, IL, USA, 1969.
2. Atmanspacher, H.; Filk, T.; Romer, R. Quantum Zeno features of bistable perception. *Biol. Cybern.* **2004**, *90*, 33–40.
3. Heisenberg W. *Philosophie: Le Manuscrit de 1942* ; Edition Seuil: Heisenberg, Germany, 1998.
4. Bohr, N. Physique. In *Atomique et Connaissance Humaine*; Edition Gallimard: Paris, France, 1991.
5. Bitbol, M. *Physique Quantique et Sciences Humaines*; Edition CNRS: Paris, France, 2009.
6. Deutsch, D. Quantum theory of probability and decisions. *Proc. R. Soc. Lond. A* **1999**, *455*, 3129–3137.
7. Busemeyer, J.R., Wang, Z.; Townsend, J.T. Quantum dynamics of human decision-making. *J. Math. Psychol.* **2006**, *50*, 220–241.
8. Busemeyer, J.R. Quantum information processing explanation for interaction between inferences and decisions. In *Quantum Informatics for Cognitive, Social, and Semantic Processes*, Arlington, VA, USA, Fall 2010; AAAI Press: Menlo Park, CA, USA, 2010; FS-10-08.
9. Busemeyer, J.; Lambert-Mogiliansky, A. An Exploration of Type Indeterminacy in Strategic Decision-Making. In *Quantum Interaction*; Springer: Berlin, Heidelberg, Germany, 2009; Volume 5494, 113–128.
10. Danilov, V.I.; Lambert-Mogiliansky, A. Measurable systems and behavioral sciences. *Math. Soc. Sci.* **2008**, *55*, 315–340.
11. Danilov, V.I.; Lambert-Mogiliansky, A. Expected utility under non-classical uncertainty. *Theory Decis.* **2009**, *2010*, 25–47.
12. Franco, R. The conjunction fallacy and interference effects. **2007**, arXiv:0708.3948v1. Available online: <http://arxiv.org/abs/0708.3948> (accessed on 20 April 2012).

<sup>32</sup>The elements of quantum formalism that we use are not complex, they are just novel to most readers.

<sup>33</sup>Moreover, the link between action and self-image is pretty weak. Benabou and Tirole recognize that when information concerns oneself, there is an important issue of attribution: is my choice a expression of the weakness of my character or are there situational reasons that justify it?

13. Khrennikov, A. *Ubiquitous Quantum Structure—From Psychology to Finance*; Springer: Berlin, Heidelberg, Germany, 2010.
14. Lambert-Mogiliansky, A.; Zamir, S.; Zwirn, H. Type indeterminacy—A model of the KT(Kahneman Tversky)- man. *J. Math. Psychol.* **2009**, *53*, 349–361
15. Mackey, G. *Mathematical Foundations of Quantum Mechanics*; Dover Publication Mineola: New York, NY, USA, 1963.
16. Kahneman, D.; Tversky, A. *Choices, Values and Frames*; Cambridge University Press: Cambridge, UK, 2000.
17. Eisert, J.; Wilkens, M.; Lewenstein, M. Quantum games and quantum strategies. *Phys. Rev. Lett.* **1999**, *83*, 3077–3080.
18. Benabou, R.; Tirole, J. Willpower and personal rules. *J. Pol. Econ.* **2004**, *112*, 848–886.
19. Benabou, R.; Tirole, J. Identity, morals and taboos: Beliefs as assets. *Q. J. Econ.* **2011**, *126*, 805–855.
20. Gul, F.; Pesendorfer, W. Temptation and self control. *Econometrica* **2001**, *69*, 1403–1436.
21. Gul, F.; Pesendorfer, W. Self control and the theory of consumption. *Econometrica* **2004**, *72*, 110–158.
22. Gul, F.; Pesendorfer, W. The revealed preference theory of changing tastes. *Rev. Econ. Stud.* **2005**, *72*, 429–448.
23. Fudenberg, D.; Levine, D.K. A dual self model of impulse control. *Am. Econ. Rev.* **2006**, *96*, 1449–1476.
24. Fudenberg, D.; Levine, D.K. Timing and Self-control. *Am. Econ. Rev.* **2011**, *3*, 34–68.
25. Strotz, R.H. Myopia and time inconsistency in dynamic utility maximization. *Rev. Econ. Stud.* **1956**, *23*, 165–180.
26. Peleg, B.; Yaari, M. On the existence of a consistent course of action when tastes are changing. *Rev. Econ. Stud.* **1973**, *40*, 391–401.
27. Zimbardo, P.G.; Cohen, A.; Weisenberg, M.; Dworking, L.; Firestone, I. The control of experimental pain. In *The Cognitive Control of Motivation*; Zimbardo, P.G., Ed.; Scotts Foresman: Glenview, IL, USA, 1969; 100–125.
28. Valins, S.; Ray, A.A. Effects of cognitive desensitization on avoidance behavior. *J. Personal. Soc. Psychol.* **1967**, *4*, 400–408.
29. Schater, A.; Singer, J.E. Cognitive, social, and physiological determinants of emotional states. *Psychol. Rev.* **1962**, 379–399.
30. Festinger, L.; Carlsmith, J.M. Cognitive consequence of forced compliance. *J. Abnorm. Soc. Psychol.* **1959**, *58*, 203–201.
31. Bem, D.J. Self-perception theory. *Adv. Exp. Soc. Psychol.* **1972**, *6*, 1–62.
32. Weick, K.E. Dissonance and task enhancement: A problem for compensation theory? *Organ. Behav. Hum. Perform.* **1967**, *7*, 345–350.
33. Berkovitz, L. The motivational status of cognitive consistency theorizing. In *Theory of Cognitive Consistency: A Sourcebook*; Rand McNally: Chicago, IL, USA, 1968; 303–310.
34. James, W. The physical basis of emotion. *Psychol. Rev.* **1994**, *1*, 516–529.

35. Beltrametti, E.G.; Cassinelli, G. The logic of quantum mechanics. In *Encyclopedia of Mathematics and its Applications*; Rota, G.C., Ed.; Addison-Wesley: Boston, MA, USA, 1981; Volume 15.
36. Maskin, E.S. Tirole, J. Markov perfect equilibrium. *J. Econ. Theory* **2001**, *100*, 191–219.
37. Fudenberg, D.; Tirole, J. *Game Theory*; MIT Press: Cambridge, MA, USA, 1991.
38. Ainslie, G. *The Breakdown of Will*; Cambridge University Press: Cambridge, UK, 2001.
39. Benabou, R.; Tirole, J. Self confidence and personal motivation. *Q. J. Econ.* **2002**, *117*, 871–915.
40. Bodner, R.; Prelec, D. Self-signaling and diagnostic utility in everyday decision making. *Psychol. Econ. Decis.* **2003**, *1*, 105–126.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>.)