Documentos de Trabajo

Lan Gaiak

THEORY OF CHOICE UNDER INTERNAL CONFLICT

Ritxar Arlegi
Miriam Teschl

D.T. 1208

Departamento de Economía

Ekonomia Saila

upna
Universidad
Pública de Navarra
Nafarroako
Unibertsitate Publikoa

A Theory of Choice Under Internal Conflict

Ritxar Arlegi
Department of Economics
Public University of Navarre, Spain
Email: rarlegi@unavarra.es

Miriam Teschl
Department of Economics
University of Vienna, Austria
Email: miriam.teschl@univie.ac.at

October 8, 2012

Abstract

In this paper we argue, inspired by some psychological literature, that choices are the outcome of the interplay of different, potentially conflicting motivations. We propose an axiomatic approach with two motivations, which we assume to be single-peaked over a certain given dimension. We first consider the case in which motivations are given and stable, and then introduce the possibility for motivations to change. We show first that in the no-motivation change case, certain choice behaviours that appear to be inconsistent from the standard rational choice point of view may be explained in our framework as the outcome of conflicting motivations. Afterwards, in the case of motivation change, we present two psychologically-flavoured assumptions about how motivations are influenced by choices. We show that, with some additional weak assumptions of rationality, motivation change leads to a smaller range of potentially inconsistent choices and not to a larger one as one may think. In particular, conflicts between two motivations can eventually be resolved by choosing different actions and consequently a definite and final preference for an action be revealed.

Keywords: Motivation, Pleasure, Self-Image,
Conflict, Preference Reversal, Dissonance reduction

JEL Classification: D01, D03

1 Introduction

The assumption that people behave on the basis of different motivations is common in the psychological and behavioural literature, and has been progressively accepted in economics, where narrow self-interest or selfishness was considered for a long time to be the most fundamental motivation. A number of models now exist that study people's choice behaviour when they experience different motivations (e.g. Bernheim 1994, Fehr and Schmidt 1999; Akerlof and Kranton 2000, 2010; Bénabou and Tirole 2002; Falk and Fischbacher 2006). A common feature of these models is that they often assume a utility function to exist. Individuals then act as standard utility maximisers and engage in some sort of a cost-benefit analysis - the choice that brings more gains given the different motivations will be chosen. Said differently, for such a utility function to exist, it must be based on a given, and stable *all-things considered* (ATC) preference ordering (see Baigent 1995) that is the outcome of a process of comparison of all alternatives taking into account all the different aspects the agent may consider relevant for choice.[1] This implies that all types of motivations must be commensurable and comparable, and that the individual can always establish a definite numerical trade-off among them. This process, however, may be quite difficult, conflictual or even impossible to resolve. What an ATC-ordering also implies is that the individual makes binding choices, that is, she is supposed to choose always the same alternative when confronted with the same choice problem.

However, in real life, many choices are not binding, even if the individual is confronted with the same opportunity set. Despite the individual's new year's resolution (as every year by the way) to live a healthier life and to drink only two glasses of wine every week and/or to practice sport three times a week, she may not keep with these decisions and start to drink more wine and/or to reduce the regular attendance at a sport club. Even though the individual may have signed up to go to the gym and to do some sport at least twice a week, she may, after some successful weeks, start to go less and less often to the gym. Repeated standard dictator game experiments show that despite the fact that subjects are exposed to the same decision situation, they are not always giving the same amount of money to their anonymous and changing counterpart at each round (Anderson et al. 2000, Bolton et al. 1998, Duffy and Kornienko 2010). The data show that a number of individuals are "hopping" between giving little in one round and giving more in the next one without

[1] A preference ordering is therefore a stable characteristic of a person. This is also why it is not unusual to refer to the preference ordering as the "identity" of the *economic* agent - an unchanging criterion with the help of which the person can be "identified" through time (see Davis 2003 and 2011, Kirman and Teschl 2004 on "identity"; see also Sugden 2004).

any clearly detectable structure. Some economic field experiments (Charness and Gneezy 2009; Acland and Levy 2010) have tested whether financial incentives may lead to binding decisions (or "habit formation" as they call it) and thus to a more regular and consistent (re)application of previous choices. Hence it can be observed that a certain option chosen at one time may not necessarily be chosen again at a later time, even when confronted with exactly the same opportunity set.

Our paper presents a new approach to decision making with several motivations for choice, whose satisfaction may be conflicting. That different motivations may lead to an internal conflict is an idea already raised by Selten (2001). Conflicts between different "points of view" of a person have also been discussed by e.g. Levi (1986) and Steedman and Krause (1986)[2]. In addition, we make a distinction between motivations and preferences: the starting-point of our analysis are the motivations of the agent, disregarding the possibility for them to be aggregated into an ATC-preference ordering[3]. In general, aggregating motivations into a single ordering implies that there exists a systematic way to solve every conflict, and this is precisely what we do not want to presuppose[4]. The purpose of this paper is to explore what kinds of choices are reasonable to be made in the presence of conflicting motivations. Therefore, choice is the result of the interplay of different motivations, but, unlike the ATC-ordering based models, such choice may or may not reveal an ATC preference according to the conventional consistency requirements of choice (see e.g. Sen 1971). That is, while we assume that there is a primitive, i.e. motivations, that generate choices, such choices may or may not generate preferences. In our axiomatic approach, motivations are supposed to be single-peaked over a certain dimension. Their numerical representation is purely ordinal and we impose as little structure as possible regarding comparability and possible "trade-offs" between them, avoiding in any case cardinal comparability.

We want to avoid any assumption on comparability for two reasons. One is that it has been shown that people do have difficulties in making trade-offs under several circumstances. Beattie and Barlas (2001) report that people have difficulties in making trade-offs when faced with a decision between, for example, money and numbers of life saved. This is a

[2] See also May (1954) and Kelsey (1986) for a somehow related discussion. Weber et al. (2001) discuss the difficulty of making trade-offs between different aspects for decision making.

[3] Thus we do not propose to attribute particular weights to the motivations, or do not discuss the possibility of aggregating them axiomatically into a single ordering such as in Steedman and Krause (1986).

[4] The only "economic" case where an ATC ordering exists and conflicts persist are instances in which the decision maker exhibits some preference or a bias towards the present. In such a case, an additional discounting variable is added to an existing utility function. This may be considered as an "ad hoc" solution. We would rather explain such a situation as a conflict in motivations and observe the resulting choice behaviour. We will in fact discuss an example of such a choice situation at the end of section 3.

situation that could be translated in a framework of competing motivations, such as, for example, a conflict between living a long life and wanting to be well-off. In fact, they reason that in general, decisions between non-commodities (such as health, security, pain, etc.) and commodities (objects sold in markets) or currencies (money and time) are more difficult to make than a choice between commodities and currencies[5]. Butler and Loomes (1988) show that in the context of decision under uncertainty, people do not find it easy to make "certainty equivalent" valuations of risky prospects. Such a situation, they reason, may lead to very imprecise preferences. Second, we consider the assumption of being able to compare any kind of motivation to be a very strong assumption on human behaviour. It is not evident that people are able to compare different motivations that may guide their behaviour such as, say, politeness and ambitiousness, or cleanliness and curiosity. In psychology, Schwartz (1994) for example presented the theory that people are guided by ten particular motivational value types, which, to different degrees, may come into conflict with each other. The very existence of conflict suggests that people may find it difficult to make decisions taking into account all of their motivations.

In a first part of our paper we consider choice under conflict with given and stable motivations. Later, we introduce the possibility of endogenous motivation change on the basis of two psychological principles, namely "reinforcement" and "dissonance reduction". That is, motivations will change as a consequence of choices made by the individual. By reinforcement we mean that the choice of an option increases the motivation for this choice and close options. Dissonance reduction follows a state of dissonance that is induced by conflicting choices that the agent is attempting to reduce or to eliminate by adapting her motivations.

Our main findings are twofold, namely that that inconsistencies in a standard rationality sense are narrowly linked to the choice of conflictual actions and should therefore not be considered as an anomaly. Second, under specific circumstances, it may turn out that we find a final and stable ATC-preference based on motivations after having engaged in a sequence of (maybe even seemingly irrational) choices.

The paper is structured in the following way. Section 2 presents the basic setup of our approach and is divided into two subsection, the first one explaining the structure of motivations, and the second one giving first notations, definitions and some preliminary results. Section 3 presents the results under the assumption that motivations are not changing. Section 4 then introduces motivation change and considers what happens if the individual

[5] This may suggest that the assumption of ATC-preferences over market goods is more appropriate than the assumption of ATC-preferences overactions and non-market goods.

engages in a sequence of actions. Section 5 discusses the possiblitiy of an equilibrium in our framework, that is, under which circumstances we can ensure that the choice of the decision maker is not going to change anymore; Section 6 concludes and presents further research questions.

2 Basic Setup of Choice Under Internal Conflict

2.1 The structure of motivations

We start our analysis by considering two types of motivations. These could be any two (potentially conflicting) motivations, but to give this theoretical analysis some context, we will call these two motivations *pleasure* (P) and *self-image* (S). The interpretation of pleasure is straightforward. As for self-image, we see it as a mental representation of the goals or ideals the person wants to achieve, or thinks she should achieve or even ought to achieve given the society in which she lives in. In that sense, the self-image as we understand it is very broadly defined and can be a purely individualistic concept as well as one that is strongly influenced by norms or rules of the society and what they say on the adequate behaviour of that person[6]. The set of alternatives from which a person is going to choose one option is a set of *actions*. We assume that choices are the result of the interplay of the two motivations: the pleasure they procure and the fact that certain actions will bring the person "closer " to or "further away" from her self-image. At this stage of the paper, we will analyse decision making on the basis of given and stable motivations. In a later section however, we will consider decision processes in which motivations change.

We assume that the two motivations can be represented as single-peaked orderings. This means that from the point of view of each motivation, actions can be ordered according to a certain dimension, whereby the degree of fulfillment of the motivation at stake may at first be increasing and then decreasing, or either always increasing or always decreasing. [7] We normalise the dimension over which the motivations are single peaked to the interval

[6] Although there may be other motivations that lead to conflicts, we think that pleasure and self-image as we characterise them are sufficiently general to explain a number of different situations. As mentioned above, Schwartz (1994) presents a theory of ten potentially conflicting motivational value types. These include, among others, "hedonism" and "stimulation" that can be compared to our "pleasure". Other motivation value types such as "self-direction", "benevolence" or "conformity" for example could all be subsumed under our idea of "self-image".

[7] A formal definition of single-peakedness is provided in the next section.

[0,1] and actions will thus be represented by points in the interval [0,1].[8] For example, the dimension over which both motivations are ordered could be the intensity with which a certain action is carried out. For example, suppose the dimension is “jogging” where 0 is no jogging and 1 is jogging every day for one hour. Suppose also that we consider a person who is not too fond of doing sports, but who would also like to be relatively slim and fit. The peak of the P-ordering for such a person could be, say, somewhere in the first third of that dimension (jogging with moderate speed sometimes on Sundays). The S-peak could be situated somewhere in the second half of the dimension (getting up early in the morning before going to work twice a weak in order to do go jogging for an hour). Single-peakedness here means that taking an action that is closer to the peak will provide more pleasure (self-fulfillment) than an action that is further away from the peak. “Learning Japanese” is an action that enables a person to speak Japanese, depending again on how much effort and time the person invests into this action. While to speak Japanese fluently may correspond to the person’s self-image, learning Japanese is a tedious task and the pleasure associated with it may be rather low, therefore, along the dimension “effort made to learn Japanese”, the pleasure peak would be at a rather low level of the dimension, while the self-image peak would lie at a high level of it.

A point in the dimension could also be seen as something more complex or more general than the intensity of executing a single action. For example, the person could consider a set of actions in terms of how they contribute to leading, say, a healthy life. A particular point on this dimension would then represent a particular “lifestyle” (e.g. more or less healthy lifestyle). For instance, a person, interested in healthiness could rank “lifestyles” from 0: eating fast-food, drinking alcohol, taking drugs and sleeping a few hours every day without any kind of physical activity to 1: following a perfectly-measured diet, doing sport every day for at least one hour and sleeping exactly 8 hours a day). Having a peak according to S in this case would mean that, for example, her optimal self-image consists in “being a rather reasonably healthy person but one who admits a certain minimum flexibility in habits” (this would be, say, point 0.8 over 1), and having a peak according to P could mean that the lifestyle that she actually enjoys the most consists in following a rather healthy diet, but also going out with friends and having some alcoholic drinks from time to time, and never doing sport because she simply does not like it (let us say 0.4 over 1). Hence, what we claim is that our model applies as far as it is possible to evaluate actions according to one particular dimension (for example, being a healthy person) and rank them in a simple

[8] For this it is needed that the set of actions has an upper and a lower bound. Moreover, for simplicity we will also assume that such set of values is continuous. Given the kinds of interpretations we adopt for the dimension both are plausible conditions.

single-peaked fashion in terms of both, P and S according to that dimension.

The assumption of single-peakedness of the motivations, together with the assumption that they are ordered over the same dimension are an important simplification of the model. However, we have found it a fruitful way to understand the problem we want to analyse. It should again be pointed out that in our analysis we do not assume single-peaked preferences. What we assume is single peakedness of the two motivations for choice. The choice that will arise as a result of the decision process will not, in general, be razionalizable by means of a single peaked preference, or even by means of a complete and transitive preference ordering. The chosen option will simply be the outcome of the interplay of the two single-peaked motivations. This means that our assumptions are conceptually weaker than assuming single peakedness of preferences because each of the two motivations belong to a more elementary and primary conceptual category[9].

2.2 Notation, Definitions and some Preliminary Results

At this stage, we will introduce some basic notation and definitions. X will denote the set of feasible alternatives that can be chosen by the individual. The alternatives that we are considering here are *actions*. Choosing an action will depend on particular motivations. These motivations are represented as single-peaked *orderings* defined over X.[10]

We will denote the class of single-peaked orderings that can be defined over X by Σ. In our setting, we will work with two single-peaked orderings defined over X, one representing the *pleasure* motivation, P, and the other representing the *self-image* motivation, S, of a person. In general we assume that $P \neq S$, otherwise the problem becomes vacuous. The peaks of P and S will be denoted by $\hat{P}$ and $(\hat{S})$ respectively. The fact that $P \neq S$ means that the individual experiences some kind of internal conflict.

Given the two motivations, the decision maker (DM from now on) makes decisions based on how *pleasant* they are and how *self-fulfilling* (or *S-fulfilling*). Moreover, we assume that the *status quo* (henceforth SQ), that is, the point of the interval [0,1] that is currently chosen

[9] Ryan and Deci (2000), psychologists known in economics because of their research on "intrinsic" and "extrinsic" motivations, for example write: "To be motivated means *to be moved* to do something. A person who feels no impetus or inspriation to act is thus characterized as unmotivated, whereas someone who is energized or activated toward an end is considered motivated." This overlaps well with our understanding of motivations, which we try to capture with the assumption of single-peakedness.

[10] Let $\succeq$ be an ordering (transitive and complete binary relation) defined over X, and let $\succ$ be its asymmetric factor. Let L be a linear ordering defined on X. We say that $\succeq$ is single-peaked with peak $\hat{\succeq}$ if, for all $x \in X$ such that $x \neq \hat{\succeq}$, $\hat{\succeq} \succ x$, and for all y such that $\hat{\succeq} L y L x$ or $x L y L \hat{\succeq}$, $y \succ x$ (see, as basic technical references, Black 1958, Inada 1969, or Moulin 1980).

by the DM, is the reference according to which an action is considered to procure pleasure or displeasure and to produce S-fulfilment or S-non-fulfilment. The idea is similar to Kahneman and Tversky's (1979) notions of gains and losses with respect to a reference point. Anything that is above the current level of pleasure (S-fulfillment) is pleasant (S-fuflilling), anything below it is unpleasant (S-non-fulfilling). Formally, since $\hat{P}$ and $\hat{S}$ are single-peaked binary relation defined over a connected subset of $\mathbb{R}$, we can define a real numerical function for P and S, ℓ_P and ℓ_S respectively, which represent those binary relations. It should be remembered that ℓ_P and ℓ_S are not *utility functions* because P and S are not *preference* relations. They should be interpreted as numerical orderings of the degree of pleasure and S-fulfillment respectively. Then, given a SQ, an action x is formally said to be pleasant (S-fulfilling) if $\ell_P(x) \geq \ell_P(SQ)$ ($\ell_S(x) \geq \ell_S(SQ)$). If $\ell_P(SQ) > \ell_P(x)$ ($\ell_S(SQ) > \ell_S(x)$), the action is considered to be unpleasant (S-non-fulfilling). At this point we should notice that the interpretation of ℓ_P and ℓ_S is purely ordinal, that is, they are unique up to any monotonic transformation.

The definitions that follow below allow us to classify the actions according to the extent they "fulfill" the two motivations of choice. There will be actions that satisfy both, the pleasure and self-image motivation, satisfy either of them or none of them. We will call these actions respectively *A, B, C or D-type of actions.* Formally:

Definition 2.1. *An action x is an A-type action if $\ell_P(x) \geq \ell_P(SQ)$ and $\ell_S(x) \geq \ell_S(SQ)$ with at least one strict inequality.*

What this means is that an A-type action is either strictly better than the SQ in both aspects (P and S) or strictly better in one of them without being worse in the other. Given that this type of action satisfies both motivations, we say that even though the DM is experiencing internal conflict due to the fact that the two peaks of the motivations do not coincide, she will however not be faced with a *conflictual action* as such when she is choosing to carry out such an action.

Definition 2.2. *An action x is a B-type action if $\ell_P(x) > \ell_P(SQ)$ and $\ell_S(x) < \ell_S(SQ)$.*

A B-type action is an action that procures strictly more pleasure than the SQ, but at the same time is not S-fulfilling. A person who chooses a B-type action not only experiences internal conflict because peaks are not overlapping, but also chooses a *conflictual action* in the sense that it provides more pleasure but less self-fulfillment.

Definition 2.3. *An action x is a C-type action if $\ell_P(x) < \ell_P(SQ)$ and $\ell_S(x) > \ell_S(SQ)$.*

A C-type action is an action that is unpleasant, but helps the person to satisfy his self-image, i.e. it is strictly S-fulfilling. Here again, the person will be faced with the choice of a *conflictual action* as only one of the motivations is satisfied.

Definition 2.4. *An action x, with $x \neq SQ$, is a D-type action if $\ell_P(x) \leq \ell_P(SQ)$ and $\ell_S(x) \leq \ell_S(SQ)$.*

A D-type action is an action that, not being the status quo, does not strictly improve in any of the aspects. In that sense we interprete that, if taken, will not be a *conflictual action.*

At this point, we assume that all types of action may be chosen. It should also be noted that even if A-type actions are available, this does not preclude that any other type of action are chosen. For example, it may be that the DM considers that the gain in one of the motivations is so important to her, even if she had to face a loss in the fulfillment of the other motivation, that she rather chooses a conflictual B- or C-type action over an A-type action that gives some gain in fulfillment in both dimensions.

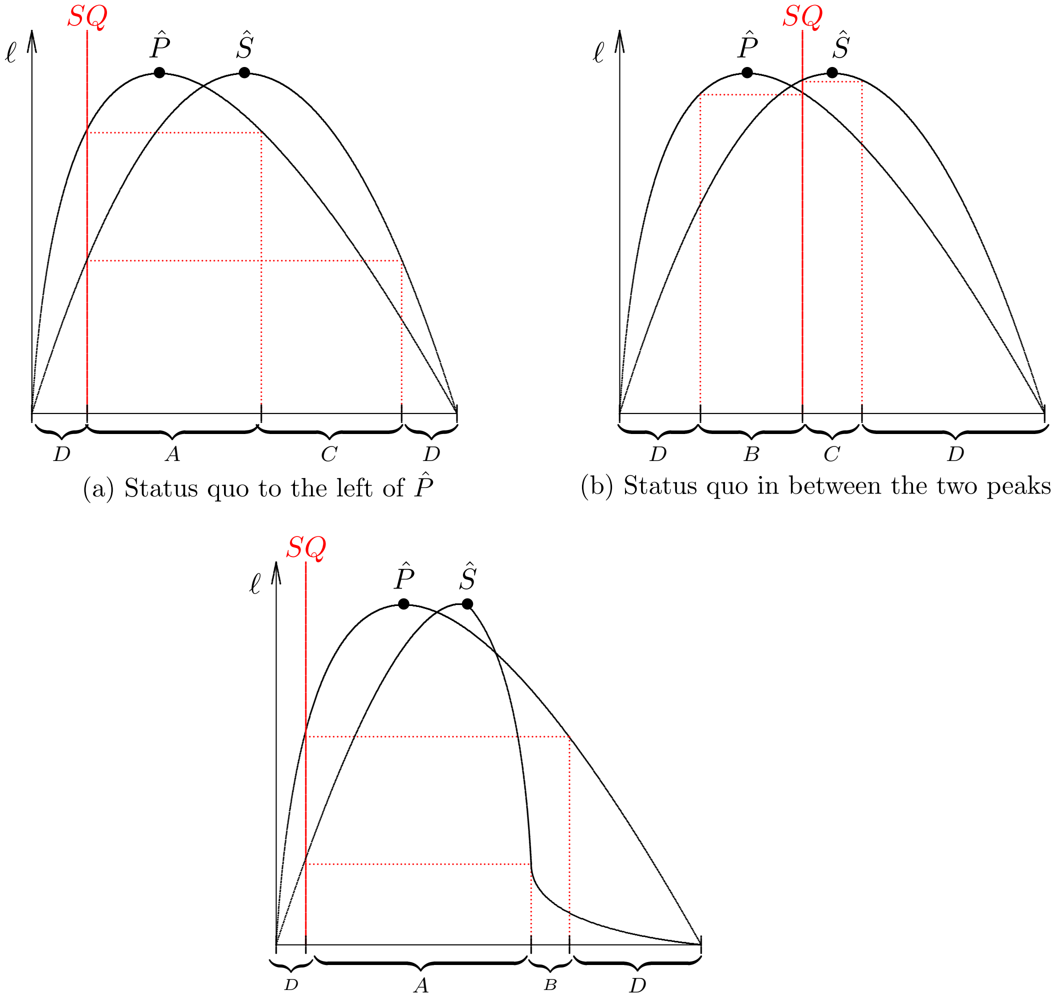
The four kinds of actions are graphically represented in Figure 1 by means of three different examples. Without loss of generality, all figures are drawn such that the peak $\hat{P}$ lies to the left of the peak of $\hat{S}$.[11] As one can see, where the different types of actions lie will depend on the current status quo. In Figure 1(a) and (c) we have represented the different set of actions for the case where the status quo SQ lies to the left of the peak of P and in Figure 1(b) the case where the status quo lies in the middle of the two peaks. Figure 1(c) represents a case in which A-type actions are followed by B-type actions and not by C-type actions as in Figure 1(a).

It is easy to check that the four types of actions plus the SQ make a complete classification of the set X. That is, every action which is not the status quo belongs to one, and only one of those types. Furthermore, since P and S are of a purely ordinal nature, the classifications of actions are independent of monotonic transformations of the two orderings. [12]

Given the definitions above, we can present the following preliminar results concerning the compatibility of certain types of actions, which will be used for later results.

[11] For simplicity we will maintain this assumption throughout the paper. If $\hat{S}$ were to the left of $\hat{P}$ it would be enough with interpreting whatever dimension we are considering in inverse terms.

[12] Graphically, this means that the classification of the actions (and all results and conclusions of our model) are not affected by "stretching" or "shrinking" vertically the representations of P and/or S, provided that pairs of alternatives with equal values of ℓ_P (ℓ_S) maintain this equality after the transformation and that every horizontal cutting of the functions lies on the same corresponding values in [0,1] before and after the transformation.



(a) Status quo to the left of $\hat{P}$

(b) Status quo in between the two peaks

(c) Status quo to the left of the peaks with B-type actions

Figure 1: Different types of actions (curly brackets indicate the respective range of actions)

Result 2.5. *If $SQ < \hat{P}$ or $SQ > \hat{S}$, then B and C-type actions are incompatible*

Proof. Given that both P and S are single-peaked orderings and that $SQ < \hat{P} < \hat{S}$ we have that for every $x < SQ$, $\ell_P(x) < \ell_P(SQ)$ and $\ell_S(x) < \ell_S(SQ)$, that is, x is a D-type action. On the other hand, given the single-peakedness of both P and S and given that $\hat{P} < \hat{S}$ we have that for all $x \in (SQ, \hat{P}]$ x is an A-type action. For the rest of the proof, we can therefore concentrate on the set of actions $\Delta = \{x : x > \hat{P}\}$.

We will distinguish two cases: Case 1: there exists $x^* \in (\hat{P}, \hat{S}]$ such that $\ell_P(x^*) = \ell_P(SQ)$. Case 2: there does not exists $x^* \in (\hat{P}, \hat{S}]$ such that $\ell_P(x^*) = \ell_P(SQ)$.

In the first case for every $x \in (\hat{P}, x^*]$ we have that $\ell_P(x) \geq \ell_P(SQ)$ and $\ell_S(x) > \ell_S(SQ)$, that is, x is an A-type action. For every $x > x^*$ we have that $\ell_P(x) < \ell_P(SQ)$ which means that x is either a C-type action or a D-type action, but never a B-type action (which requires $\ell_P(x) > \ell_P(SQ)$) due to the single-peakedness of P and that $\hat{P} < x^*$.

For case 2 we can assume that there exist y, z, $y \neq SQ$, $z \neq SQ$ such that $\ell_P(y) = \ell_P(SQ)$ and $\ell_S(z) = \ell_S(SQ)$. If those two points would not exist (e.g. imagine "truncated" P or S orderings such that according to the SQ all actions are pleasant and S-fulfilling to the right of the SQ, or that $\hat{S}$ is at point 1), then it is not difficult to see that there are no B-type actions if such a z does not exist, and no C-type actions if such a y does not exist, or there are neither B or C-type actions if neither such a z nor such a y exist. In any case the impossibility for both types of actions to co-exist would be proven. Therefore, take such a pair of y and z actions. If $y < z$ we have that, due to the single-peakedness of both P and S, on the one hand that for every $x \in (\hat{P}, y]$, $\ell_P(x) \geq \ell_P(SQ)$ and on the other hand that for every $x \in (\hat{P}, z)$, $\ell_S(x) \geq \ell_S(SQ)$. Therefore, given that $y < z$, and by single-peakedness, for every $x \in (\hat{P}, y]$ $\ell_P(x) \geq \ell_P(SQ)$ and $\ell_S(x) > \ell_S(SQ)$, that is, for every $x \in (\hat{P}, y]$, x is an A-type action, while for every $x > y$ $\ell_P(x) < \ell_P(SQ)$, being x either a C-type or a D-type action, but never a B-type action. In the case where $z < y$ it can be proved analogously that either B or D-type actions are at the right of z, but never a C-type action. Finally, when $z = y$ we have that only D-type actions arise at the right of $z(y)$.

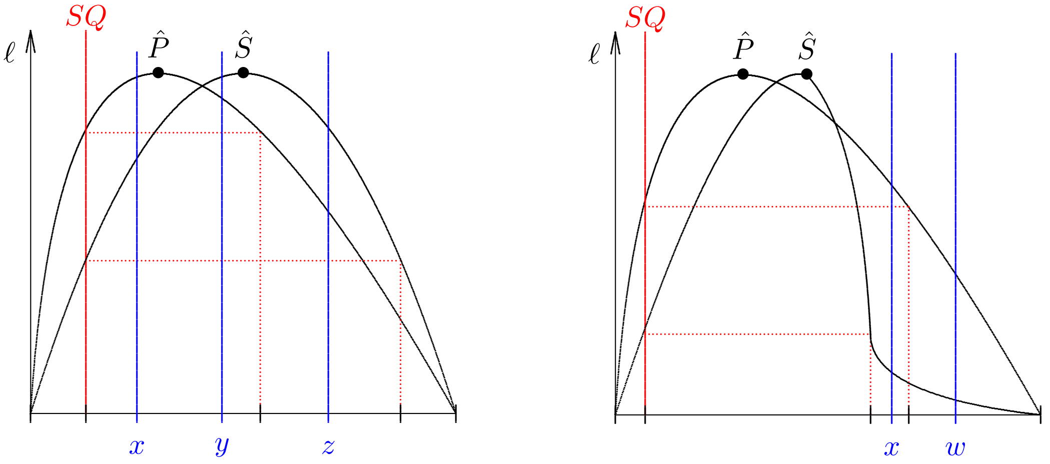
To prove that when $SQ > \hat{S}$ then B and C-type actions are incompatible can be made by following analogous steps. □

Result 2.6. *If $\hat{P} < SQ < \hat{S}$ then there are no A-type actions.*

Proof. By definition x is an A-type if $\ell_P(x) \geq \ell_P(SQ)$ and $\ell_S(x) \geq \ell_P(SQ)$ with at least one strict inequality. Given that $\hat{P} < SQ$ and by the single-peakedness of P we have that for $\ell_P(x)$ to be greater or equal than $\ell_P(SQ)$ it is a necessary condition that $x \leq SQ$. But now, given the single-peakedness of S and given that $SQ < \hat{S}$ we have that, for every x, if $x \leq SQ$, then $x < \hat{S}$, and therefore $\ell_S(x) < \ell_S(SQ)$, that is, $\ell_P(x) \geq \ell_P(SQ)$ and $\ell_S(x) \geq \ell_P(SQ)$ together are impossible, and therefore there are no A-type actions.

□

Results 2.5 and 2.6 highlight the importance of the status quo and the assumption of single-peakedness of the two motivations for choice when considering the kinds of conflicts an individual might face. In particular, Result 2.6 depicts a situation in which the individual is at a point in between the peaks of the pleasure ordering and her self-image ordering, and will therefore necessarily have to be dithering between two conflictual types of actions,



(a) x and y are A-type actions; z is a C-type action

(b) x is a B-type action; w is a D-type action

Figure 2: Choosing an action

namely the B and C-type actions. In other words, any action that she may take different to the status quo involves a conflict between the two motivations for choice because it means moving in one or the other direction. For example, imagine the case of somebody who follows a not strict enough diet for what she considers to be the best one, but too strict for what she enjoys eating; or a professor who devotes a too small portion of her time to teaching for what she considers to be the best standard, but too much for what she likes teaching. Under such circumstances, moving from the status quo implies, in any case, moving away from one of the two references for choice, that is, our DM either makes a stricter diet (teaches more) obtaining less pleasure or relaxes her diet (teaches less) moving away from her optimal self-image.

In the situation displayed by Result 2.5, the status quo is at a relative “extreme” end of the spectrum of actions, and this implies that only one of the conflictual types of actions is present in addition to non-conflictual A-type actions. Now, the agent has room to improve according to both motivations, for example by choosing x as in Figure 2(a). She could also improve according to both motivations by choosing an action that is even *further* to the right than her pleasure-peak, as action y in that same figure. However, from a certain moment onwards, for any actions that are even further to the right of the dimension the individual has to pay a price in terms of the satisfaction of at least one of the two motivations and A-options will not any longer be available. It can be the case that the action is less pleasant but more fulfilling in terms of the self-image (C-type) (action z in Figure 2(a)). It can also be the case that the action is not S-fulfilling but becomes worthy in terms of increased pleasure (B-type) as action x in Figure 2(b), or even that an action becomes a D-type as action w in Figure 2(b), which happens when actions become unpleasant and non-S-fulfilling at the same time. Due to the single-peakedness of the motivations, it is clear that, if the SQ is to the left of the peak of P, there can be no situation in which actions that are C-type (B-type) are followed by actions that are B-type (C-type) further to the right of the dimension.

As a corollary of Results 2.5 and 2.6 we therefore have that, if A-type actions exist, then B and C-type actions are incompatible. This means that the existence of A-type actions does not exlude the possibility of being faced with the potential choice of a conflictual action, but that the kind of conflictual action she faces is either of a B- or a C-type. This is because, as we said above, in our model, even if A-type actions are available, we take it that the DM does not necessarily have to choose an A-type action. We assume that it is perfectly possible that she chooses a B- or a C-type action if she values highly enough the gain in a particular motivation, even if such an action involves a loss in the other motivation, and even if there exists other alternatives that involve, for example, small gains in both motivations.

3 Choice with Stable Motivations

One of the aims of the paper is to show that certain kinds of behaviour that would be irrational from a standard rationality point of view can be explained as plausible in our theory. In our approach, instead of aiming to find an ATC-preference ordering adapted to the behaviour we want to explain, we start by assuming that there are subsets of "admissible" and "non-admissible" alternatives according to certain axioms. This fits with a "satisficing" approach to decision making á la Simon (Simon 1955, 1956) but not with a maximising approach where only the best option will be chosen. This, obviously, means putting less structure on solving the problem, and consequently allows for more "irrationalities" to be explained. However, our main contribution is to show that, even under the weak requirements of our setting, interesting results can be obtained, linking inconsistencies with the existence of conflict between motivations and in particular with the choice of conflictual actions.

Formally, let $C : 2^X \times X \Rightarrow X$ be a *choice function* that assigns to an opportunity set and a status quo a unique choice. For any $K \subseteq X$, $x = C(K, z)$ will be read as "x is the choice from set K when the status quo is z". We assume that, if K is the opportunity set, the status quo always belongs to it. Moreover, in some cases it will be relevant for the development of our model specifying the time at which a choice is given, in such a case we will denote by $x = C_t(K, z)$ a situation where "x is the choice from set K at time t when the status quo is z". The admissibility of a choice will be determined on the basis of certain conditions regarding the interplay of the two motivations and the status quo. We start by presenting the following simple axiom:

Axiom 3.1. Status Quo Low Monotonicity (SQLM): *$\forall x \in X$, $x \neq SQ$, $\forall K \subseteq X$, if $\ell_p(x) \leq \ell_p(SQ)$ and $\ell_s(x) \leq \ell_s(SQ)$ then $x \neq C(K, SQ)$.*

The axiom says that if an action x, when compared with the status-quo, is worse with respect to the two motivations under consideration, then it will never be chosen. Given the classification of actions above, it is clear that SQLM directly makes any D-type actions inadmissible.

One of our main objectives is to better understand, by means of our formal tools, why certain apparent inconsistencies arise. A straightforward example of inconsistency according to the standard theory is *preference reversal*, that is, a situation where the DM chooses x over y at one time, but another time she chooses y over x from the same opportunity set. We start with a narrow definition of this kind of phenomenon in connection with the status quo, which we call more appropriately status quo dependent *choice* reversal:

Definition 3.2. *We say that* status quo dependent choice reversal *holds when, for some $x, y \in X$ and some $K \subseteq X$, $y = C(K, x)$ and $x = C(K, y)$.*

Status quo dependent choice reversal displays a typical situation of "dithering" between two alternative actions or lifestyles, each appealing for different reasons. The dithering means, to use Jon Elster's terminology that "The grass is always greener on the other side of the fence" (Elster 1989, p.9). Once the DM has settled for one action, he longs to go back to his previous situation, but once he is there again, he bemoans not to be realising the other action. This dithering is the consequence of the changing SQ and the corresponding re-evaluation of existing actions. It is reminiscent of Buridan's ass dithering, only that in the original example, the ass does not know how to rank tow identical haystacks, whereas here, one haystack would have golden looking hay, but be further away than the other haystack whose hay would have not such an appetising colour, but is much closer. In our case, Buridan's ass would first choose one haystack, but once it has chosen it, the ass is not sure whether the other one would not have been better and chooses that one the next time. Like in the original Buridan's ass example, where the ass is unable to choose which of two identical haystacks is better, it does not make sense, as Sen (1973) explains, to assume that the ass is indifferent between the two options: Indifference is not the cause for the dithering, but the fact that the ass does not know how to rank the two haystacks vis-à-vis of each other. This is a qualitatively different information of the ass' choice behaviour than indifference. Another evidence that dithering is not equal to indifference is provided by Beattie and Barlas (2001), who point to experiments that have shown that people have more difficulties to decide between two alternatives who are similar to each other, rather than between those that are very different. [13]

More worldly situations of "dithering" are frequent. Some smokers reduce their number of cigarettes they smoke per day (or even quit smoking) for a certain period of time, only to smoke more (or start smoking again) some months later. Other people face a motivation problem relative to sport: for some time they are motivated to go, say, jogging three times a week, only to lose their motivation again and to go jogging only on Sundays.

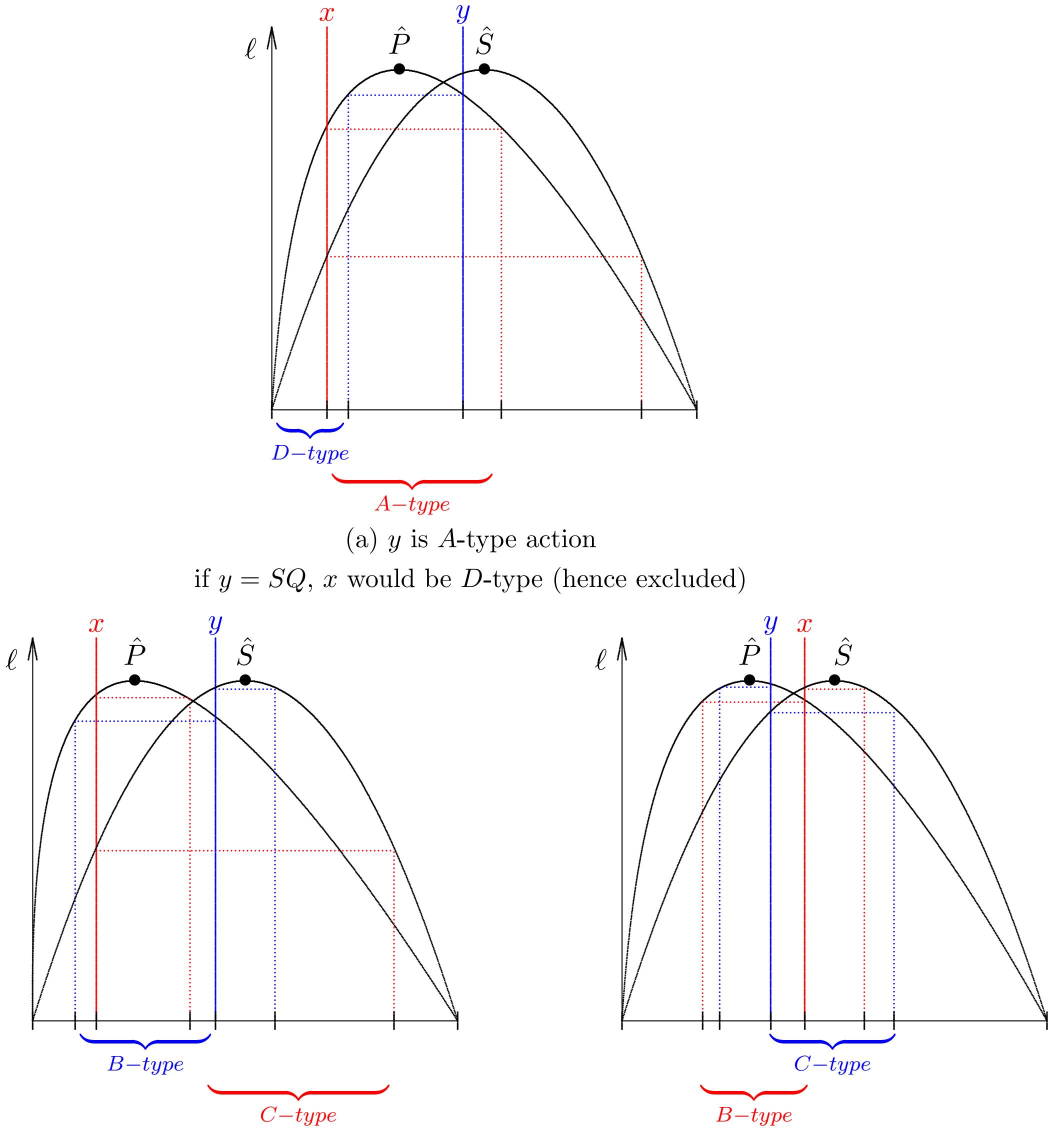
Result 3.3. *Assume that SQLM holds. If for some $x, y \in X$, and for some $K \subseteq X$, $y = C(K, x)$ and $x = C(K, y)$ and thus* status quo dependent choice reversal *holds, then y is either a B or a C-type action with respect to x and x is a B or a C-type action with respect to y. Moreover, if y is a B-type (C-type) action and x, the previous SQ, is chosen, x will be a C-type (B-type) action.*

Proof. Assume that $y = C(K, x)$ and $x = C(K, y)$. First, y is not a D-type by SQLM. On the other hand, y cannot be an A-type because then x would become a D-type and, under SQLM, choice reversal will not arise –see figure 3(a) in which we assume that $K = X$. Hence, for choice reversal to arise, whatever set K is, the SQ has to be an admissible action once y has been chosen, and this only happens when y is a B or C-type action. With this we prove that choice reversal is only possible when y is either a B or a C-type action with respect to x. That x is either a B or a C-type action with respect to y is proved analogously.

Now we prove that both, B or C-type actions may lead to preference reversal. We will do it by means of two examples. Consider, for example, $K = X$. In figure 3(b), y is a C-type action. When y becomes the SQ, action x, the former SQ becomes a B-type action and thus is an admissible action. In figure 3(c) we depict a situation in which y is a B-type action. When y becomes the SQ, x is a C-type action and thus admissible.

The fact that if y is a B-type (C-type) action when $x = SQ$ then, x, when chosen again, is a C-type (B-type) action can easily be proven by the very defnition of B- and C-type actions.

[13] Actually, as it is well-known, even by assuming indifferences, inconsistencies cannot be avoided, as is illustrated by numerous famous examples, such as Luce's (1956) coffee/sugar example or Tversky's (1972) Paris/Rome example.



(a) y is A-type action
if $y = SQ$, x would be D-type (hence excluded)

(b) y is a C-type action;
if $y = SQ$, x is B-type

(c) y is a B-type action;
if $y = SQ$, x is C-type

Figure 3: Status-quo dependent choice reversal with B-type and C-type actions

□

This result suggests that choice reversal is strongly related with the choice of conflictual actions, that is, with B or C-type actions. Choice reversal reveals a kind of “Anti-SQ-bias”: the option that one does not have is always better. Jon Elster (1989) has used this “The grass is always greener on the other side of the fence” syndrom to explain what he calls “counter-adaptive preferences”. For Elster, such a situation arises when people are observed to prefer options that are not available to them. He also calls this phenomenon "forbidden fruits taste best": people prefer options that are out of reach for them. Our model explains this "syndrome" not because options are not available, but because people choose conflictual actions, which they are not able to trade-off in any consistent way to determine which ones they would prefer. This situation implies that individuals tend to focus on the good attributes they are forgoing (better fulfilling their self-image or experiencing higher pleasure respectively) and to pay less attention to what they are currently experiencing. This motivates this pendular switching of their choices. That is, for *status- quo dependent choice reversal* to arise it is necessary that, when moving from $SQ(x)$ to y: (i) y is not an A-type (or D-type); (ii) the DM decides to put a lower weight on the loss in $P(S)$ than on the gain in $S(P)$ and (iii) the DM is willing to make the same trade-off but in the opposite direction when moving from y to x, being y the SQ, that is, to put again a lower weight on the loss in $S(P)$ than on the gain on $P(S)$.

Our notion of choice reversal is reminiscent of Tversky and Kahneman's (1991) model of *loss aversion*, where gains and losses are assessed in a two dimensional space. With their theory they can explain observed phenomena such as status-quo bias or endowment effect. Only they assume that people value losses always more than corresponding gains. Furthermore, given their assumption of comparability between the two dimensions, they establish a certain correspondence between these gains and losses that applies to every decision of the agent. We do not propose such correpondence but simply state that for a particular case of choice reversal, gains must be valued more than losses.

Obviously, *status quo dependent choice reversal* is not the only kind of inconsistency in choice that one can observe. Inconsistencies are generally those that violate some consistency condition. The consistency condition that ensures that a preference relation exists, which rationalizes the choice function is *Independence of Irrelevant Alternatives* (See, for example, Kalai, Rubinstein and Spiegler 2002; Binmore 2009). Next we propose an adaptation of that condition to our context with the aim to analyse under which circumstances it will be violated. Its violation in fact will include *status quo dependent choice reversal* as a particular case.

Definition 3.4. *Let us consider the following condition (IIA): $\forall K, T \subseteq X$, $K \subseteq T$, $\forall x \in K, T$, if $x = C_t(T, z)$ for some $z \in T$, $t \in \mathbb{N}$, then $x = C_{t+1}(K, x)$. We will say that* IIA is violated *when, for some $K, T \subseteq X$, $K \subseteq T$, and for some $t \in \mathbb{N}$, there exist $x \in K, T$, $y \in K, T$, $z \in T$, such that $x = C_t(T, z)$ and $y = C_{t+1}(K, x)$. In that case we will say that* the pair (x, y) leads to the violation of IIA.

In words, the definition above reflects a situation where the DM maker chooses a certain action x when y is available, but in the following situation where we have fewer actions but both actions are still available, the DM chooses y. Here, unlike with the *status-quo dependent choice reversal*, inconsistencies may arise even with A-type actions. Inconsistency may happen, not in relation to the SQ, but in relation to a third action that, having been rejected when the action x was chosen can be chosen when x becomes the SQ. For example, consider again Figure 2(a). To simplify, assume that X is the opportunity set in every situation, which is a particular case under the domain of IIA. Suppose action x was chosen first, even though y was also available. In this case, action x is becoming the new SQ. In a following choice, even under SQLM, it is admissible that the DM chooses y even though x is also available this leads to *IIA violation*.

Compared with *status quo dependent choice reversal*, *IIA violation* opens the door to a large range of inconsistent behaviours. Given that, it is reasonable to think about further axioms that restrict the admissibility of actions.

Axiom 3.5. Dominance (DOM): *$\forall K \subseteq X$, $\forall x, y \in K$, if $\ell_p(x) \leq \ell_p(y)$ and $\ell_s(x) \leq \ell_s(y)$ with at least one strict inequality then, for all $z \in K$, $x \neq C(K, z)$.*

DOM generalizes the SQLM axiom for any pair of actions (not only the status quo versus another one). The consequence is that an action x will never be chosen if there exists another action y such that $\ell_p(y) \geq \ell_p(x)$ and $\ell_s(y) \geq \ell_s(x)$ with at least one inequality. As a consequence of SQLM we excluded D-type actions only. With DOM, besides D-type actions, many more actions are excluded. In particular, it is easy to see that, if the whole set of actions X is available, or if K contains actions between the two peaks, DOM constrains the range of admissible actions to them. This leads us to the following result:

Result 3.6. *Assume that DOM holds. If a pair of actions (x, y) leads to IIA violation, then y can be either a B or C-type action only.*

Proof.

Assume that $x = C_t(T, z)$ for some $z = SQ$, $t \in \mathbb{N}$ and $T \subseteq X$, and let $I = \{w \in X : \hat{P} < w < \hat{S}\}$. That is, I contains all the actions that are between the two peaks. Then two cases arise: (i) $T \cap I \neq \emptyset$. Then, by DOM, $x \in I$, and by Result 2.6, there are no A-type actions at time $t+1$. Thus, any action y that is chosen at $t+1$ will never be an A-type action. By DOM, y can never be a D-type action either. Therefore, it can only be a B-type or a C-type action. (ii) $T \cap I = \emptyset$. Then, T consists of elements at the left of $\hat{P}$ and/or at the right of $\hat{S}$, that is, $T = A \cup B$, where $A = \{a \in I$ such that $a \leq \hat{P}\}$ and $B = \{b \in I$ such that $b \geq \hat{S}\}$. Given the single peakedness of P and S, there are three possibilities concerning the set of admissible actions that DOM imposes: First, such a set consists of $max_{\ell_p}(A)$ if $max_{\ell_p}(A) \geq max_{\ell_p}(B)$ and $max_{\ell_s}(A) \geq max_{\ell_s}(B)$ with at least one strict inequality. Second, such a set consists of $max_{\ell_p}(B)$ if $max_{\ell_p}(B) \geq max_{\ell_p}(A)$ and $max_{\ell_s}(B) \geq max_{\ell_s}(A)$ with at least one strict inequality. And third, the set of admissible actions is $\{max_{\ell_p}(A), max_{\ell_p}(B)\}$ otherwise. In the first case, by DOM, for any $K \subseteq T$ such $max_{\ell_p}(A) \in K$, $C_{t+1}(K, max_{\ell_p}(A)) = max_{\ell_p}(A)$, and IIA cannot be violated. The second case is analogous. In the third case, by DOM, being $x = C_t(T, z)$, necessarily $x \in \{max_{\ell_p}(A), max_{\ell_s}(B)\}$. Without loss of generality we will assume that $x = max_{\ell_p}(A)$. Consider now $K \subseteq T$ such that $x \in K$. By DOM we know that, if $C_{t+1}(K, x) \neq x$, necessarily $C_{t+1}(K, x) \in B$. By hypothesis, we know that, for every $b \in B$ it is not true that $b \geq max_{\ell_p}(A)$ and $b \geq max_{\ell_s}(A)$ with at least one strict inequality. That is, there are no A-type actions in B, and thus no violation of IIA can hold with an A-type action. A direct consequence of DOM is also that IIA violation is impossible with D-type actions.

What we have proved is that, under DOM, the IIA violation cannot be with y being a D or A-type action. What we still have to prove is that, under DOM, such a violation can result with y being either a B or a C-type action. It is not difficult (but we think not necessary) to show examples where x is either an A, B or C-type action and y is either B or C-type and where the IIA is violated (obviously x can never be a D-type action by DOM). □

Result 3.6 connects once again inconsistencies more closely with conflict. If DOM holds then for IIA to be violated, at the latest from the the second choice onwards, it must be either a B or C-type action. That is, even if the first action is of an A-type, IIA is violated only if a conflictual action (improving in one motivation but renouncing in the other) is chosen next.

As we mentioned above, *IIA violation* implies a larger set of possible inconsistencies than *status-quo dependent preference reversal.* If we only impose SQLM, then *IIA violation* arises even with A-type actions and this even repeatedly so. The strengthening of SQLM to DOM means that, though the initially chosen option can be an A-type, the violation of IIA arises with the choice of B or C-type actions only.

Next we propose an additional axiom that will be used to show that the availability of the choice of some non-conflictual A-type actions is a special situation, which may have an influence on whether the DM chooses consistently or not. Let us first consider a new definition:

Definition 3.7. *Let $x = SQ$. We say that an action y is engaging with respect to x if either $x < \hat{P} < y$ or $y < \hat{S} < x$.*

By an engaging action we mean that the DM takes a decision that moves her new SQ that previously was either to the left of the P-peak or to the right of the S-peak beyond the peak of at least one motivation. This means that an engaging action can be of either type, as long as the peak of at least one motivation lies in between the original and the new SQ. An action is engaging if, in some way, the DM has overcome or "vanquished" a hurdle, where the hurdle is the peak of at least one motivation. We can now introduce our new axiom.

Axiom 3.8. Commitment (COM)*: $\forall K \subseteq X$, $\forall x, x', y \in X$ such that $\ell_p(x') \geq \ell_p(x)$ and $\ell_s(x') \geq \ell_s(x)$ and such that x' is engaging with respect to x. If $y \neq C_t(K, x)$ then $y \neq C_{t+1}(K, x')$ for all $t \in \mathbb{N}$.*

Axiom COM, which, like SQLM and DOM, is of purely ordinal nature, means that the DM takes a committing decision to which she will stick in the future and from which she will not deviate. We understand action x' to be committing if it fulfills two conditions: on the one hand it is *engaging*, that is the action goes beyond one peak, and on the other hand it is an A-type action, which means that the DM experiences an improvement in both dimensions. We can imagine such an action to be a kind of "cold turkey" action. As is well known, to stop smoking "cold turkey", for example, would mean that the DM stops smoking from one day to the other; he or she does not gradually stop smoking by smoking less every day. Assume that the SQ of such a smoker is to the left of the two peaks, which means that the DM currently smokes too much for what he or she likes smoking, but also for how much he or she would ideally like to smoke (assume for example a truncated S-motiation where the peak of S is close to 1, i.e. ideally, according to his or her self-image, the DM would like to stop smoking). In such a situation, a committing decision to stop smoking altoghether not only is an A-type action, but a decision that shifts the SQ to the very other side of the dimension and thus is engaging, meaning that at least one of the peaks lies between the old SQ and the new one.

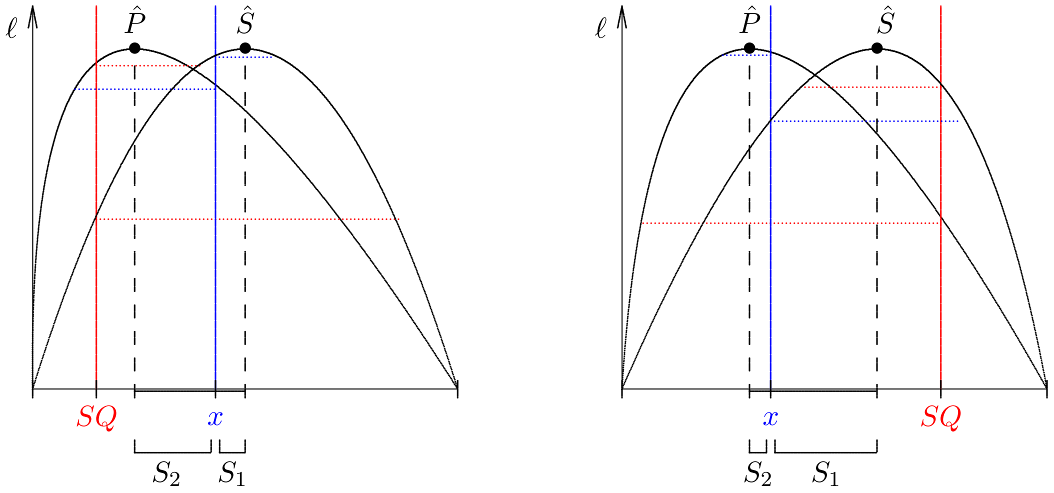
Committing decisions are also reminiscent of "precommitment", as Elster (2000) for example understands it, or "commitment" as Schelling (2006) would call it. Both mean by this that the DM benefits from restricting his or her set of opportunities. For Elster (2000), "precommitment" is a form of rationality over time where "[a]t time 1 an individual wants to do A at time 2, but anticipates that when time 2 arrives he may or will do B unless prevented from doing so." (p. 5). Such a person would look out for a (pre)commitment device in order to make sure that another choice than A is impossible or at least less likely.

In our case, a committing decision means that at time 2, the DM will not deviate and choose something different, that is, he or she will commit with the choice first made. In this sense, the DM voluntarily restricts her opportunity set to a particular action. Why should she do so? Our approach gives some structure to the reasons why the DM commits or activates some kind of commitment device.[14] The DM knows that when she chooses an A-type action that is engaging at time 1, from that moment onwards, she will only be faced with B, C or even D-type actions, which necessarily involve some kind of loss if she were to deviate from the new SQ, i.e. the action to which she committed. At time 1, the DM can commit to an engaging action because the prospect of the dithering between different conflictual actions that necessarily involves gains and losses at the same time where she to deviate from the new SQ at time 2 is not considered worthy taking given the "positive" experience of having chosen an A-type action, through which she was able to improve in both dimensions. At time 2, this evaluation is confirmed: the prospect of having to choose between different actions that necessarily involve losses is valued less than the renewed choice of the SQ. In such a situation, the DM prefers to stick with the original gains, rather than to suffer additional losses. In other words, the DM can commit to an action at time 2 if she values losses more than gains, that is, when losses affect her more than gains. This is of course the opposite statement we made with respect to *status quo dependent choice reversal*: the latter only arises if the DM values gains always more than losses. Here, a person can commit and thus acts consistently if she values losses always more than gains.[15] Hence, in some sense contrary to Tversky and Kahneman's (1991) result, which explains some well-known examples of preference reversal saying that people value losses always more than corresponding gains, in our case, valuing losses more than gains makes people act consistently.

Another way of explaining the choice of a committing action is to say that the DM is not myopic, as one may argue that a person has to be if she engages in *status quo dependent*

[14] A commitment device is usually something that helps a person to carry out the decision taken at time 1. A person who decided not to drink alcohol at home for example, may have thrown or given away her bottles of alcohol in order not to be tempted to renege on her decision. People who have the tendency to overspend money may pay their money into bank accounts, which penalises them if they withdraw more than once a year. For other examples, see e.g. Elster (2000).

[15] Someone may object and say that a person cannot commit if she is a non-exponential discounter for example. As we mentioned in the introduction, non-exponential discounting is a standard case of "preference" reversal in the economic literature. However, as we have also indicated in the introduction, models with non-exponential discounting usually assume a utility function and then add the discount factor. Here, remember, we have no preferences and thus no utility function. If there are no preferences, then even if a person is a non-exponential discounter, it is not clear that she would deviate from her committed action. This is because if she is not able to compare different motivations with each other, the fact that she discounts non-exponentially does not change that situation. Furthermore, the person would have to discount gains and losses and not simply gains at different moments of time. This may also complicate the decision problem.



(a) x is C-type action

(b) x is a B-type action

Figure 4: Violation of IIA with B and C-type actions

choice reversal for example. Rather, the DM may decide that the one-time gain in both dimensions is better than any future gains and losses she would face if she deviated from the committed choice. Given that, she makes a form of one-time-for-all decision, which would imply a habitual action from then onwards, an action whose advantages and disadvantages she no longer evaluates.

It then follows that if COM holds in addition to DOM we have the following result.

Result 3.9. *Assume that DOM and COM hold. If a pair $x, y \in K \subseteq T$ leads to a violation of IIA, that is $x = C_t(T, z)$ for some $z \in T$ and $y = C_{t+1}(K, x)$, and $T \cap [\hat{P}, \hat{S}] \neq \emptyset$, then x and y can be only B or C-type actions.*

Proof.

Assume that x is an A-type action, then by Result 2.6, the SQ is not in between the two peaks. Also, by hypothesis we know that there exist elements in between the two peaks, and by DOM we know that x must then be in between the two peaks. Therefore x is engaging with respect to SQ, and by DOM this prevents any other alternative y that was available at the first choice to be chosen, thus making the violation of IIA impossible. By DOM, we know that x cannot be a D-type action. Therefore x must be either a B or a C-type action. From this point on the proof of Result 3.6 can be replicated to prove that y is either a B or a C-type action.

Now we have to prove that, being x and y either B or C-type actions, the violation of IIA is possible under the axioms. We will prove this by showing two figures that display the admissibility of four possible cases, namely, that both actions are B-type actions, that both actions are C-type, that x is B-type and y is C-type, and that x is C-type and y is B type. For simplicity we assume in all cases that $K = T = X$, but it is not difficult to find examples where $K \subset T \subset X$. In figure 4(a) we see that the chosen x is a C-type action. Later, any y that may be chosen in the segment S_1 will be a C-type action, and any y' that may be chosen in the segment S_2 will be a B-type action. None of the two described sequences $(x; y' \in \{S_2\})$ or $(x; y \in \{S_1\})$ violate DOM or COM. In figure 4(b), x is a B-type action. Later, any y that may be chosen in the segment S_1 will be a C-type action, and any y' that may be chosen in the segment S_2 will be a B-type action. Again, none of the two described sequences $(x; y' \in \{S_2\})$ or $(x; y \in \{S_1\})$ violate DOM or COM.

□

In the case of *status quo dependent choice reversal* we have restricted the set of acceptable situations by SQLM, which ruled out D-type actions. The result we obtained was that *status quo dependent choice reversal* was only associated with B and C-type actions, that is, with the choice of conflictual actions.

Results 3.6 and 3.9 come to a conclusion that is, in a general sense, similar to that of Result 3.3. We find a link between the choice of conflictual actions and inconsistency. Since the *IIA violation* is a more general kind of inconsistency than *status quo dependent choice reversal* it is logical that, for arriving at the same kind of result we need more conditions than just SQLM. If we impose DOM, IIA is will be violated when at the latest the second choice is either a B or C-type action. If the opportunity set is such that there are also actions in between the two peaks to choose from, we also impose COM, we find an even stronger link between Result 3.9 and Result 3.3 in the sense that inconsistencies arise as in the latter result with B- or C-type actions only. This is because COM, together with DOM and the assumption that there are choices in between the two peaks, rules out any kind of inconsistency once the DM chooses an A-type action.[16]

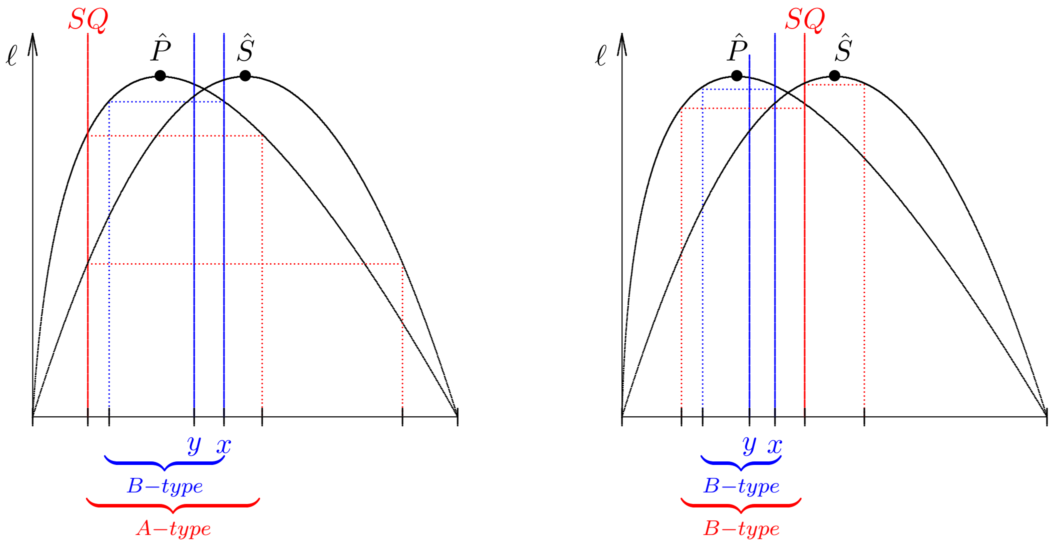
In the following, we will turn to two explicit and well-known examples of motivation conflict and will set out how these conflicts can be explained in our framework. The first example in particular is aimed at clarifying the meaning of Result 3.9 providing an interpretation of Ainslie's party example.

[16] A commiting choice of a particular action is in fact equivalent to a situation in which the DM maximises some existing underlying preference that is a convex combination of the two motivations. This does not mean that the DM has, strictly speaking, solved the conflict, which would only be the case when the two peaks coincided, but she or he was able to assign, for example, appropriate weights to the two motivations that reflected the best trade-off for the DM.

Example 3.10. **Ainslie's party example:**

Ainslie's (1992) account of a person's decision to go to a party is a particular example for motivation conflict. A person's objective (self-image) is to pass an exam on Monday. However, the person is invited to a party on Sunday, and he would like a lot to go to this party. But if he goes to that party, he knows that he may risk to fail in the exam, especially if he stays for a long time and has several drinks. So what he decides is to go to the party for a short while, but to leave the party early enough, at say 10pm, in order to have enough sleep before the exam. The crucial moment is when 10pm is approaching because the question is, whether he is going to act according to his previous decision and to go home (and thus to come closer to his self-image by renouncing to some pleasure) or will he stay at the party (and thus increase his pleasure at the cost of his self-image achievement)? Our analysis above gives us a clear answer to that question. Suppose that the dimension according to which he ranks his pleasure and his self-image is to be a more or less *responsible person*. Along this dimension, 0 indicates the least responsible person, 1 the most responsible person. Suppose that currently he is not a very responsible person: he goes to many parties, drinks quite a lot, goes out often, etc., which gives him some pleasure, but actually makes him also feel sick and tired. Moreover this life is very far from his self-image goals. His SQ thus lies to the left of the responsibility dimension, and left of the P-peak. However, this person has now decided to change his lifestyle and to become a more "serious" person - hence his intention is to pass the exam on Monday. There is quite a wide range of "improvement" possible for that person in terms of both, pleasure and self-image. He knows that he would be able to enjoy life and to achieve his self-image as long as he is able to live a balanced life, which means that he studies enough and restricts pleasant activities to a few hours per day. DOM already restricts the set of decisions about the intended leaving time. Assume that the person takes the decision to go to the party until 10pm (action x) that is an engaging A-type action given the current SQ, which is to the left of the peak of P. According to our analysis above, when 10pm approaches, by COM the person will not change his mind by choosing to stay longer (action y), but will leave the party without experiencing any kind of further conflict. He may have arranged a commitment device such as having ordered a taxi to arrive at the party at 10pm which would make it harder for him to renege on his decision. He thus would not be revealing any kind of inconsistency.

However, suppose that the person is already a rather responsible person and his SQ lies in the middle of the two peaks as in figure 5(b). Then, going to the party and even if he leaves at 10pm is an act that goes against his self-image, but it would enhance his pleasure. Under such circumstances, this person may well reconsider his previous decision to leave



(a) The *irresponsible* person's decision problem (by COM y is excluded)

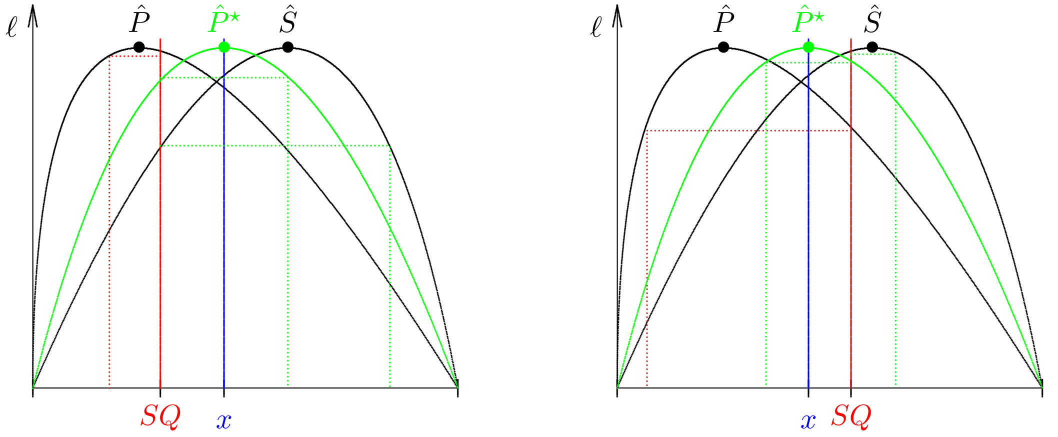
(b) The *responsible* person's decision problem (here y is a permissible choice)

Figure 5: Ainslie's party example

at 10pm once 10pm approaches and stay longer at the party (action y in figure 5(b)). He may therefore act inconsistently according to the standard theory. DOM however restricts the set of possible inconsistent actions the person could choose from. The reason for this inconsistency is that the choice the person has made even before going to the party was already a conflictual B-type action and the conflict is continuing to matter for the person once he is at the party when he has to implement his previous decision. This analysis suggests that the choice of a conflictual action may induce further choices of conflictual actions. This example therefore suggests, even though it seems counterintuitive, that the person who is currently more responsible has more chances to act inconsistently than a person who is currently less responsible. In fact, being a responsible person, he may not think that it is useful to organise a taxi to pick him up at the party and thus opens the door to revise his decision as the party goes along.

Example 3.11. **Extrinsic and intrinsic motivation:**

The next example discusses the potential conflict between intrinsic and extrinsic motivations for effort, and in particular the possible "crowding out" of intrinsic motivation due to an increase in extrinsic motivation. Extrinsic motivation means that a person does some work primarily because of something external to that work, such as a reward or a recognition. Intrinsic motivation on the other hand is the motivation that drives or pushes



(a) Intrinsic and extrinsic motivations reinforce each other

(b) Crowding-out of intrinsic motivation

Figure 6: Intrinsic and extrinsic motivation example

the individual from “whitin” to do certain actions, such as the pleasure or satisfaction the person obtains for the sake of doing the job (see e.g. Deci and Ryan 1985; Frey 1994). Translated into our context, we can say the following. Suppose that *effort* is the dimension that we now consider: 0 means no effort, and 1 means maximum effort. We assume that for more effort, the individual is paid a higher wage. Let us now interpret our two peaks. Suppose that P represents the pleasure or “satisfaction” a person receives from being paid a particular wage for a given effort level and thus represents the individual’s extrinsic motivation. The individual’s associated pleasure of putting in more effort and gaining more is at first increasing, but then decreasing. This is reminiscent of the more standard assumption of a decreasing marginal utility of money: even if at higher levels of effort the individual would be paid more money, there is a point where the satisfaction of receiving more money for more effort declines. S on the other hand represents the satisfaction the person receives from doing the work itself, with an analogous interpretation and thus represents the person’s intrinsic motivation. Suppose then, that in our example, the peak of P is to the left of the peak of S, which would mean that for lower levels of effort, the monetary reward is more important than for higher levels of effort. In other words, from a certain level of effort onwards (the peak of P), the individual would be willing to increase his effort for the satisfaction he receives from doing the job for the sake of the job, but would not be willing to increase his effort just considering the corresponding increase in income.

Suppose now that the current SQ lies in between the two peaks (he is working too much as for the extrinsic motivation but too little as for the intrinsic one). The employer now wants to incite the person to invest more effort into doing his job and proposes a wage increase. A wage increase could be represented as an external shift of the P ordering to the right, with the new peak being $\hat{P^{*}}$ (the green peak in Figure 6), all other things remaining the same: for low effort levels this implies that to achieve the same satisfaction as before, the individual had to work more. At higher effort levels, he would be incited to work more as he receives more satisfaction from the wage increase than before.

We can differentiate between two situations. In the first situation, represented in figure 6(a), the wage increase shifts the peak of P to the right so that the new peak of P, is now to the right of the SQ. This can happen in particular when the SQ of that person, i.e. his current effort level, was more strongly determined by his wage than by the activity itself. In such a case, the change of P offers a new set of A-type actions to the person (indicated with green dotted lines): the SQ now lies to the left of both peaks, $\hat{P^{*}}$ and $\hat{S}$. In this situation, given that everything to the left of SQ has now become a D-type action, by SQLM the person is indeed going to increase his effort (for example, by choosing a new A-type action which increases both, his satisfaction from being paid and his pleasure from doing this activity). In our figure, as an example, we assumed that the person is actually choosing a new effort level that gives him the highest pleasure in terms of money he will be paid, while at the same time experiencing more intrinsic motivation too. But it could of course have been a different action too. The worst that can happen for the employer is that this person is not going to increase his effort, but remains at the SQ.

In the second case (figure 6(b)), suppose that the SQ of the person was already closer to $\hat{S}$ than to $\hat{P}$. Here, an increase in wage may result in shifting the peak of P to the right, but not enough so that $\hat{P^{*}}$ still remains to the left of the SQ. In such a case, the wage increase does not solve the motivation conflict that the person has between intrinsic and extrinsic motivation. He is still caught in a situation of choosing between conflictual B or C-type actions in addition to simply remaining at the SQ. In such a situation, it may even happen that the person decides to decrease his effort. While he would lose some of his pleasure from doing the job in itself, he will be more satisfied with earning more money than before for a lower effort level.

4 Endogenous Motivation Change

A crucial aspect of our model is that engaging in an action of whatever type is not innocuous regarding future decisions. As we have seen in the previous section, this is because taking an action leads to a change of the status quo, and consequently the whole partition of actions among different types changes as well.

In addition to this change in the reference point, we introduce next the idea that doing a particular action will have an effect on people's motivations, and consequently on people's choices. That is, motivations will change endogenously by making choices.[17] Our assumptions about how choices affect motivations are expressed by means of axioms with a psychological connotation. In the following we will concentrate on two specific kinds of psychological experiences, but this does not exclude the possibility of other psychological influences on motivations with different effects as the ones we propose. Moreover, for reasons of simplicity, we will assume troughout this section and the next one that every action will be available at any moment, hence the opportunity set will be X in any case.

The first psychological axiom is called *reinforcement*. In our context it means that making a particular action will increase the pleasure a person receives from doing it. That is, doing a particular action will tend to augment the motivation for doing it in terms of its pleasantness. The other source of motivation change will be called *dissonance reduction*. In our context this means that by choosing an action whose level of self-image satisfaction declines with respect to the status quo, the person experiences a form of dissonance that he wants to reduce. This is translated into a change of ordering S in such a way such that the individual is going to render it more coherent with his choice.

A short comment is in order at this point. It could be claimed that reinforcement and dissonance reduction are two psychological phenomena that can be associated with pleasure and self-image, but not necessarily with other conflicting motivations. Thus, unlike the previous sections, the theory developed in this section could not be applied to other motivations. It may be true that reinforcement and dissonance redution apply to certain motivations only, but certainly not to pleasure or self-image alone. Reinforcement may be associated more generally with the enjoyment of an activity where the person increases satisfaction by engaging in that activity (listening to music, playing an instrument, etc.). This would be similar to the idea of habit formation as Stigler and Becker (1977) for example. But reinforcement may also apply to motivations such as courage or curiosity, where courage and curiosity increase with an accumulated history of actions done on the basis of those motivations. Dissonance reduction does also not need to be associated with changes in self-image only, but can be of various qualities and origins, such as changes in the level of altruism, adherence to social norm, execution of duty, etc. Hence, by changing the meaning of the P and S-motivation, we are able to apply the theory to a number of different settings. For example, there could be conflicts between creativity or curiousity on the one hand and obedience or honoring parents on the other, ambitiousness could conflict with a

[17] Note that the endogenous motivation change we are going to introduce is different to the extrinsic movement of the P peak we discussed in the second example of the previous section.

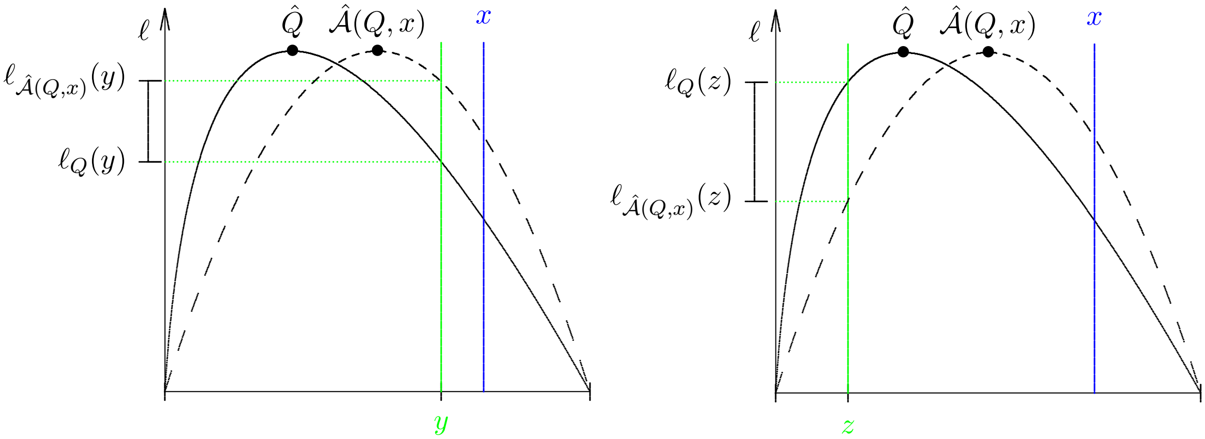
particular social order, etc. Schwartz's (1994) motivational value structure to which we also referred in Section 2 shows that motivational value types such as "security" or "power" can conflict with individual freedoms, equality, environmental concerns, etc., motivations that he subsumes under "universalism" and "self-direction" value types (p. 37). In summary, what is clear is that we propose a theory that assumes particular pairs of conflicting motivations to which it makes sense to apply reinforcement and dissonance reduction.

For a precise definition of reinforcement and dissonance reduction we first need the define what is an *adaptation function.* It formalises the idea that a single-peaked ordering *moves towards* a point x in X. According to the definition below, when the ordering moves in one direction (for example in the direction of x, x being at the right of the peak as in figure 7), the peak of the new ordering will be placed somewhere in between the old peak and x; everything at the right of the new peak $\hat{\mathcal{A}}(Q,x)$ will have a higher value - see figure 7(a); and everything at the left of the old peak $\hat{Q}$ will have a smaller value - see figure 7(b). The definition is general enough in two important senses: on the one hand it allows the new peak to be anywhere in between the old peak and x, and on the other hand it does not specify how the value of the options between the old peak and x is affected. Formally:

Definition 4.1. Adaptation Function *An adaptation function is a function $\mathcal{A} : \Sigma \times X \rightarrow \Sigma$ such that, for any $Q \in \Sigma$, for any $x \in X$,*

- *If $x > \hat{Q}$ then $x > \hat{\mathcal{A}}(Q,x) > \hat{Q}$; for all $y > \hat{\mathcal{A}}(Q,x)$ we have that $\ell_{\mathcal{A}(Q,x)}(y) > \ell_Q(y)$ and for all $z < \hat{Q}$ we have that $\ell_{\mathcal{A}(Q,x)}(z) < \ell_Q(z)$*
- *If $x < \hat{Q}$ then $x < \hat{\mathcal{A}}(Q,x) < \hat{Q}$; for all $y < \hat{\mathcal{A}}(Q,x)$ we have that $\ell_{\mathcal{A}(Q,x)}(y) > \ell_Q(y)$ and for all $z > \hat{Q}$ we have that $\ell_{\mathcal{A}(Q,x)}(z) < \ell_Q(z)$*

In our model, the adaptation function will provide a tool to express the idea that both P and S may change as a consequence of the actions taken by the agent due to reinforcement and dissonance reduction. Our idea of adaptation is reminiscent of the idea of "aspiration adaptation" (Selten, 1998) according to which, like in our case, taking actions produces a certain adaptation in aspirations and goals. We will denote by $P'(x)$ (respectively $S'(x)$) the situation in which it is the pleasure ordering P (the self-image ordering S) that adapts to a new pleasure ordering (self-image ordering) following the chosen action x. Also, whenever P (S) becomes a new ordering $P'(x)$ ($S'(x)$) we will say that P (S) *adapts* to x; that it *moves towards* x; or in some cases, that it *moves* left or right, depending on whether the ordering adapts to some action which is to the left or to the right respectively. With this in mind, we can state next our two axioms.



(a) $y > \hat{Q}$ and $\ell_{\hat{\mathcal{A}}(Q,x)}(y) > \ell_Q(y)$

(b) $z < \hat{Q}$ and $\ell_{\hat{\mathcal{A}}(Q,x)}(z) < \ell_Q(z)$

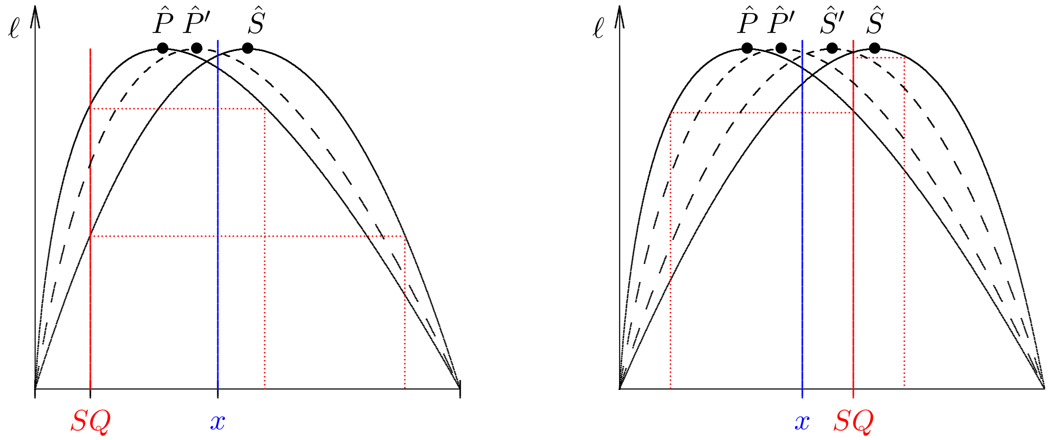
Figure 7: Adaptation function with action x to the right of $\hat{Q}$ ($x > \hat{Q}$)

Axiom 4.2. Reinforcement (RF) *For all $x, z \in X$, such that $x = C(X, z)$, P adapts to x resulting in a new pleasure ordering $P'(x)$ if and only if $x \neq \hat{P}$.*

Being an *if and only if* condition, this axiom has two readings. On the one hand it says that whenever the agent takes an action x different to the SQ, this has an effect on the pleasure ordering, which adapts by moving towards it. On the other hand, though there may be many factors that affect the pleasure motivations, the axiom reflects our interest to concentrate on the influence of choice on P. The intuitive idea of RF is that any action x will induce P to change such that the current action will become more pleasant, but also *close* ones will do. For example, if x is to the right of $\hat{P}$, all those that are further to the right than x and some actions that are in between x and $\hat{P}$ will increase their level of pleasure. Meanwhile, all the actions that are *further away* from the taken one will become less pleasant (all those that are further left than $\hat{P}$ – see Figure 8(a)).

Jon Elster (1989, chapter 9; 2007, chapter 16) has also discussed this experience. For Elster, like for us, reinforcement is not necessarily a conscious goal of action. In his view, reinforcement is the fact that when an action has pleasant consequences, it increases the probability of engaging in it. This is not exactly what we mean. Our idea is that doing an action increases its pleasantness, but we give no indication as to whether the person *therefore* is going to choose it with a bigger probability or not, given that the person will still need to decide between two (competing) motivations.

Axiom 4.3. Dissonance Reduction (DR) *For all $x, z \in X$, then S adapts to x resulting into a new self-image ordering $S'(x)$ if and only if $x = C(X, z)$ and $\ell_S(x) < \ell_s(SQ)$.*



(a) Reinforcement: $\hat{P}$ moves towards x (**CCM1**)

(b) Dissonance reduction: $\hat{P}$ and $\hat{S}$ move towards x (**CCM2**)

Figure 8: Two psychological axioms

In this case, S will change if and ony if the agent takes an action x which fulfills her self image less than the current status quo (see figure 8(b)). The idea is that the individual will have to "rationalize"[18] ex post the choice of her action in order to reduce the dissonance that was created by engaging in an action that distances her from her self image. DR incorporates the findings of psychological research on cognitive dissonance following Festinger (1957), which predicts that people will be changing some of their cognitions if two or more of them are in conflict with each other. By changing the S-ordering, the person accomodates the self-image to make it more consistent with the chosen action and thus to reinstall some "consonance". Indeed, by changing S, $\ell_S(x)$ increases for any chosen action and thus offsets to some extent the loss she would has experienced by acting in opposition to her self-image.

It can be noted that, if we assume SQLM, we know that D-type actions (unpleasant and non-S-fulfilling) are ruled out. Consequently, it is not possible that somebody takes an action which is non-S-fulfilling unless it is pleasant, so that, there is only room for DR if x is pleasant.

In the case of S-adaptation induced by DR, we could potentially also have a situation in which the self-image changes so much that it "overshoots" the P (or even the $P'(x)$ ordering). Suppose for example that the status quo currently lies to the right of $\hat{S}$ and the person can choose from a set of A and B-type actions. If, for whatever reason the

[18] The term "rationalize" is used in economics as well as in psychology. In psychology it describes a defense mechanism used by a person to make something appear logical and more consistent with one's actions and this is what we mean in this context.

person chooses a B-type action that is close to 0, $\hat{P}$ will move towards x, but because $\ell_s(x) < \ell_S(SQ)$, the person will experience dissonance, and in order to reduce it, will change S. Given that $\hat{S}$ will move towards x, it may well go beyond $\hat{P}$ ($\hat{P'}(x)$) and overshoot it.

Another remarkable fact is that RF and DR do not exclude the possibility that an action does not affect neither P nor S. In particular, there is one case where an action does not lead to any change in motivations, and this is when $x = \hat{P}$ and, at the same time, it is an S-fulfilling action. This observation leads to our next result.

Result 4.4. *Assume that RF holds, then any sequence of at least two different actions leads to motivation change.*

Proof. For the first action not to affect either P or S it has to be such that $x = \hat{P}$ and self-image fulfilling. If the second action y is *different* to x this means that $x \neq \hat{P}$ and by RF P changes. □

This result is speaking for itself. The only time when no motivation change happens is if the DM is engaging in an A-type action that procures the highest pleasure and is self-image fulfilling. In any other case, the choice of actions will necessarily trigger motivation change.

If we assume our two psychological axioms and SQLM, P can move either left, right or does not move, and so can S. In total, we therefore have nine possible *combined changes in motivations* (**CCMs**). However, if we impose RF, DR and also DOM, then we have our next result that restricts the set of possible **CCMs** to those where the peaks either approach or do not move.

Result 4.5. *Assume that RF, DR and DOM are fulfilled, then, for any action x, the four following combined changes in motivations (*__CCMs__*), and only those, are possible.*

- ***CCM1: Self-image peak approaching***. *We say that an action x leads to a* self-image peak approaching CCM *if it makes P moving right while S does not move.*

- ***CCM2: Dissonant self-image peak approaching***. *We say that an action x leads to a* dissonant self-image peak approaching CCM *if it makes P moving right and S moving left.*

- ***CCM3: Dissonant pleasure peak approaching***. *We say that an action x leads to a* dissonant pleasure peak approaching CCM *if it makes S moving left while P does not move.*

- ***CCM4: No change in motivations***. *We say that an action x leads to* no change in motivations *if it induces neither S or P to change.*

Proof. As we said above, in general there are nine possible **CCMs**. The three combinations where P moves left are only possible by RF if x is to the left of $\hat{P}$, but this is ruled out by DOM. Similarly, among the remaining possibilities, the two where S moves right are only possible by DR if x is to the right of S, which is again ruled out by DOM.

Now, we have to prove that under RF, DR and DOM, **CCM1**, **CCM2**, **CCM3** and **CCM4** are possible. For that it is enough to imagine the following situations:

- For **CCM1**, assume that $SQ < \hat{P}$ and x is any B-type action that is in between the peaks of P and S.
- For **CCM2**, assume that $\hat{P} < SQ$ and x is any C-type action that is in between the peaks of P and S.
- For **CCM3**, assume that $\hat{P} < SQ < \hat{S}$ and $x = \hat{P}$.
- For **CCM4**, assume that $SQ < \hat{P}$ and that $x = \hat{P}$.

□

Taking into account Result 4.4, and that **CCM1**, **CCM2** and **CCM3** have all of them the effect of approaching P and S, the interpretation of Result 4.5 is neat: Under DOM, the two psychological axioms we impose (RF and DR) imply that any sequence of different actions the individual may carry out will contribute to reduce the distance between the two peaks. The two peaks of the motivations are becoming closer, reducing the set of conflictual actions.

For our next result, we continue to assume motivation change through RF and DR. However, both axioms are not needed to obtain the result, and hence are not included in the formulation of the result.

Result 4.6. *Assume that DOM is fulfilled. Then for any sequence $\{x^n\}$ of actions such that $n > 1$ and $x^i \neq x^{i+1}$ for all $i < n$, x^i is either a B-type or a C-type action for all $i > 1$.*

Proof. By DOM we know that no element of the sequence $\{x^n\}$ can be a D-type action. Also by DOM $x^i \in [\hat{P}, \hat{S}]$, and by Result 2.6 we know that no action in $[\hat{P}, \hat{S}]$ is an A-type action. □

What Result 4.6 says is that, with DOM, if the decision maker engages in a sequence of different actions, then from the second action onwards, any action is *conflictual*, that is, a B-type or a C-type action (though the first one, which is not included in the sequence

$\{x^n\}$ might be either an A, B or C-type action). If however the chosen action is always the same, that is, the SQ is chosen over and over again, then the DM, although she still experiences motivation conflict by the fact that the peaks do not overlap, is not faced with a repeated decision among conflictual actions and thus is not faced with any gains or losses in motivations.

We can now make a global relationship among the implications of all the results obtained in this section up to this point. Result 4.4 shows that, under DOM, choosing two different actions leads necessarily to motivation change, and Result 4.5 shows that, again under DOM, such a motivation change reduces the distance between the two peaks. Furthermore, according to Result 4.6 we will have a sequence of conflictual actions (B or C-type actions except maybe for the initial choice). In the previous sections we have presented some results that show that inconsistencies in choice only arise when conflictual actions are chosen. As the two peaks approach, the range of conflictual choices shrinks. Though one could have thought that allowing for motivations to change would lead to more possible inconsistencies, our model provides the interesting (and unpredicted) conclusion that if we allow for motivations to change, then the potential range for inconsistent choices is going to be reduced.

At this point, we may wonder about the possibility of full convergence of the P- and S-peaks, that is, about a sequence of actions that makes both peaks coincide. Given how the adaptation function is defined, an action x induces the peak of a motivation to approach x, but not to fully coincide with x. A consequence of this is that, if we want both peaks to fully coincide we have to think about the problem *in the limit*. That is, we have to think about infinite sequences of actions that lead both peaks to converge in the limit.[19] However, even in terms of infinite sequences of actions, it may not necessarily be the case that they lead to full convergence of the peaks. Actually, we can distinguish two forms of convergence.

1. *Full convergence in the limit*: The (infinite) sequence of actions $\{x^n\}$ makes both peaks, $\hat{P}$ and $\hat{S}$ coincide in the limit. Formally, $\lim_{n\to\infty}\{P'^n(x^n)\} = \lim_{n\to\infty}\{S'^n(x^n)\}$. This represents a situation where the DM is involved in a continuous sequence of actions that changes motivations with the result of a full convergence of both peaks and

[19] As a matter of fact, it is possible to change the definition of the adaptation function in a way such that allows for $P'(x) = x$ (or $S'(x) = x$). This would not affect substantially the general conclusions of the model and would allow to think about full convergence without the need of assuming that the sequence of actions is infinite. However, given its meaning, it seems to us more natural to define the adaptation function as we have done it.

the progressive disappearance of conflict in the limit when both peaks coincide. That is, this case represents a final resolution of the conflict.

2. *Partial convergence in the limit*: The (infinite) sequence $\{x^n\}$ does not lead to a final coincidence of $\hat{P}$ and $\hat{S}$ in the limit. Formally, $\lim_{n\to\infty}\{P'^n(x^n)\} \neq \lim_{n\to\infty}\{S'^n(x^n)\}$. This represents a situation where the decision maker engages in a continuous sequence of actions, but which, in the limit, does not lead to a convergence of the peaks to the same point.

In general, the two types of convergence defined above can be the result of the first three types of combined changes of motivations (**CCMs**) - **CCM4** obviously does not lead to any convergence. Nevertheless, there are some restricted relationships that can be proven. As a matter of example, consider *full convergence in the limit.* This is possible with **CCM1** or **CCM2** only or with any combination of **CCM1**, **CCM2** and **CCM3**, but not with **CCM3** only. The convergence in the case of **CCM1** only happens if the sequence of actions in between the peaks is such that $x_i < x_{i+1}$ for every i and converges to $\hat{S}$, changing $\hat{P}$ such that $\lim_{n\to\infty}\{P'^n(x^n)\} = \hat{S}$. Such a behaviour would correspond to an individual who, having a strong self-image reference, has been able to make coincide her pleasure motivation to her self-image by progressive sacrifices of pleasure (C-type actions). For instance, a smoker who finally enjoys being a non-smoking person by means of giving up smoking little by little.

A particular example of partial convergence in the limit would be a situation where there is a limit in the agent's motivations' mutability, which at a certain point becomes infinitesimal. Then choices can continuously be made in the gap between the peaks, but it could be perfectly the case that $\lim_{n\to\infty}\{P'^n(x^n)\} < \lim_{n\to\infty}\{S'^n(x^n)\}$.

The example above is not the only possibility for partial convergence in the limit. We could think about a situation in which, from a certain moment onwards, the taken action is always the same. That is, the DM decides to remain at the SQ and not to change her decision any longer. Since choosing the SQ does not induce dissonance reduction, S is not going to change any more, and for that reason the convergence in the limit will never be complete.[20] This may happen if the DM has found a sufficiently convenient choice when taking into account the potential benefits and losses of changing to a more pleasant and less S-fulfilling action or to a more S-fulfilling but less pleasant action. In this case there always remains a gap between the two peaks. But in general, nothing in the model prevents

[20] The only situation where choosing continuously the same action could lead to full convergence of the peaks would be if such action is the peak of S.

the DM to choose, eventually, another action. In other words, there are no elements in the model to ensure that such sufficiently convenient choice is "definite" or "final". However, if COM holds, and the DM is choosing an engaging A-type action, then she will not further deviate from this action. The P-peak will therefore converge to the final SQ, but the S-peak will remain where it is. In such a situation, we have a final partial convergence because there will always remain a gap between the two peaks.

The observation about full or partial convergence may also give some interesting insights into welfare assessments. Rembember that we have to differentiate between two types of situations: one is that the DM experiences a general kind of internal conflict simply by the fact that the peaks are not overlapping. The other is when the DM has to make a choice between conflictual actions (that is B- and C-type actions) and the DM knows that in that case she will be faced with losing in terms of one motivation to gain in terms of another. Continously engaging in conflictual actions (choosing B- or C-type actions) has the advantage that peaks may eventually converge (full convergence) and that in the limit, the DM has resolved her conflict, and will not be faced with conflictual choices any longer. The advantage of choosing a committing action is that the DM gains at first in both dimensions and then chooses consistently the SQ from then onwards without facing any further choices of conflictual actions, but continuous to live with a general kind of internal conflict due to the fact that the peaks will not converge. Thus, if an agent is committing to a particular action, it is as if the DM considered it to be better to be consistent from a certain point onwards rather than to be faced with the continuing choice of conflictual actions, even with the prospect of achieving a final resolution of the conflict because over time peaks may converge. Said differently, when COM holds, the DM, prefers to choose one action from a restricted opportunity set over choosing different actions from a larger opportunity set. This is consistent with the understanding of commitment as explained by Elster (2000) or Schelling (2006) for example.[21]

[21] Elster (2000, p. 21) writes: "The idea of precommitment is often linked to that of second-order desires. Suppose a person wants to quit drinking, but finds himself torn between his desire to drink and his desire for all the things that drinking prevents him from doing. This conflict does not necessarily generate a second-order desire not to have the desire to drink. In general, when we desire two incompatible things we decide which desire is the more important and act on it." Our COM axiom does give precise conditions under which the DM is able to make such a decision given internal conflict of motivations. If the DM makes the choice of a committing action, then, as Elster suggests, it does not mean that the conflict as such has disappeared, which in our context means that the peaks may not overlap.

5 Equilibrium

In this section we consider one more issue, namely whether the above described convergence processes can lead to an *equilibrium.* In line with standard choice theory, we define an equilibrium as a situation where, when the opportunity set does not change, we can ensure that the decision maker will not change her action any longer.

Definition 5.1. *We will say that $x \in X$ becomes an equilibrium at date t if $x = C_{t-1}(X, w)$ for some $w \neq x$ and we can ensure that $x = C_j(X, x)$ for all $j \in \mathbb{N}, j \geq t$.*

The definition above not only describes an equilibrium as a choice that is not going to change, but it also makes explicit the date at which an equilibrium will be reached. On the other hand, for the definition to apply, it is not enough to observe that the same choice is continuously taken, but we also need to be sure that something in the model prevents any other choice to be taken. The next results show how the assumptions of SQLM and DOM affect the existence of an equilibrium and under which configurations of the peaks and the status quo it can be achieved.[22] All the results in this section assume that motivations can change, that is RF and DR hold. However, the results also apply if one assumes that motivations are fixed.

Result 5.2. *Assume that SQLM holds. If there exists $x \in X$ that becomes an equilibrium at time t, then, at time t, $\hat{P} = \hat{S}$ and $x = \hat{P} = \hat{S}$.*

Proof.

Assume that x becomes an equilibrium at a certain time t. We have two possibilities: either $x = \hat{P}$ or $x \neq \hat{P}$. If $x = \hat{P}$ then unless $\hat{P} = \hat{S}$, $\hat{P} + \epsilon$ is not ruled out by SQLM given that $l_s(\hat{P} + \epsilon) > l_s(\hat{P})$, and we cannot ensure that x is going to be chosen thereafter. Therefore x is not an equilibrium. Thus, for $x = \hat{P}$ to be an equilibrium at t it is necessary that $\hat{P} = \hat{S}$. On the other hand, if $x \neq \hat{P}$, then either $l_p(x + \epsilon) > l_p(x)$ or $l_p(x - \epsilon) > l_p(x)$. Again, SQML does not prevent $x + \epsilon$ $(x - \epsilon)$ to be chosen and x is not an equilibrium at t.

□

What the result says in words is that, under SQLM, for x to be an equilibrium, the peaks must coincide and x must be the choice of the two coinciding peaks. This is a situation where there is no further conflict between the motivations and there is an unambiguous best

[22] For the sake of fluency we will say that "x is an equilibrium" when there exists $t \in \mathbb{N}$ such that x becomes an equilibrium at t.

choice, which one could interpret *as if* the DM were maximizing some existing underlying preference.

In the next result we show the consequences of imposing DOM, which, as explained in previous sections, is a stronger version of SQLM.

Result 5.3. *Assume that DOM holds. Then, there exists $x \in X$ that becomes an equilibrium at time t if and only if $\hat{P} = \hat{S}$ at time t. Moreover, for all $w \in X$, $x = C_t(X, w) = \hat{P} = \hat{S}$.*

Proof.

DOM is a stronger axiom than SQLM, therefore, in order to prove that x is an equilibrium at time t only if $\hat{P} = \hat{S}$ at time t we can reproduce the proof of Result 5.2.

Now, we will prove that, if $\hat{P} = \hat{S}$ at time t, then there exists an equilibrium at time t and that, in particular, for all $w \in X$, $x = C_t(X, w)$ is an equilibrium. Consider $x = C_t(X, w) = \hat{P} = \hat{S}$. Note that $\forall z \in X$ such that $z \neq x$, $\ell_p(x) > \ell_p(z)$ and $\ell_s(x) > \ell_s(z)$, thus, by DOM, for all $z \neq x$, $z \neq C_{t+1}(X, x)$, and repeating the argument indefinitely we have that x becomes an equilibrium at time t, and that $x = \hat{P} = \hat{S}$.

□

The main difference between Results 5.2 and 5.3, is that, with SQLM, the coincidence of the peaks is a necessary but not sufficient condition for reaching an equilibrium. Even if $\hat{P} = \hat{S}$ it is not guaranteed that the equilibrium will be reached because SQLM does not prevent the DM to choose a different action (for example the DM may repeatedly choose the SQ which does not coincide with the peak of P and S). In this case SQLM leaves always open the possibility for another choice. DOM however ensures that the coincidence of the peaks is both a necessary and a sufficient condition for the equilibrium to be reached. Moreover, when peaks coincide, DOM guarantees that wherever the SQ is, the equilibrium will be reached at time $t + 1$. This leads to the interesting observation that although the situation where peaks coincide can be interpreted *as if* there are some well defined preferences (at least in what concerns the best alternative), SQLM is sufficiently weak to allow for conventional inconsistencies such as *IIA violation* to arise. For example, the decision maker could take a sequence of different A-type actions approaching the coinciding peaks and would thus violate IIA. This cannot happen under DOM. In that case the unique best option is chosen from the very beginning. A possible interpretation of this observation is that SQLM allows for some myopy or bounded rationality of the decision maker in the sense that she is only able to see which options are worse than the current SQ, but not to pairwise compare every available option and to decide which one is the best one when both peaks coincide, as implicitly assumed by DOM.

Finally, one may wonder at this point about the implications of assuming COM for the equilibrium issue. It turns out that if we also impose COM, we do not gain any particular insight concerning the possibility of equilibria to exist and their characteristics. As we know, COM alone, by its very statement, ensures that an action is an equilibrium if and only if it is an engaging A-type with respect to a SQ that is outside the interpeaks space. If we combine COM with either SQLM or DOM, we just add this equilibrium possibility to that of $\hat{P} = \hat{S}$.[23]

6 Conclusion

Our preceding analysis of choice with conflicting motivations allows us to highlight several things. First of all, we have seen that inconsistencies from the point of view of standard economic theory are the result of the choice of conflictual actions. If we assume SQLM, which is certainly a minimal assumption of rationality, *status quo dependent choice reversal* only happens because of the choice of conflictual choices. If we incorporate some additional assumptions of rationality, namely DOM or COM, then not only *status quo dependent choice reversal* but also *IIA violation* are actually the outcome of conflictual actions(B- and C-type actions). Both kinds of actions satisfy only one motivation at the expense of the other one and because of this, the DM may be torn between these two type of actions to experience satisfaction interchangeably in at least one of the two motivations.

Second, despite the intuition that motivation change may lead to more inconsistent behaviour, we find that if we impose some reasonable conditions such as DOM, motivation change leads eventually to a smaller range of possible inconsistencies. Hence, the fact that motivations are malleable and unstable may actually be necessary for the DM to become less inconsistent. Indeed, under certain circumstances, conflicts will end eventually and it can be said that the DM engages from then onwards in consistent choices according to standard rationality. We have considered two different circumstances in which the DM becomes consistent: one is by committing to an action, the other one is by continuously choosing conflicting actions, which has the effect that the peaks may converge and the conflict resolves.

[23] Since axiom COM applies to a situation where there is a sequence of two decisions, and it was used before assuming motivation change, its statement has to be refined in order to consider the effects of the motivations change because it is not clear *when* the hypothesis of the engaging A-type action should apply, that is, either before motivations change (time t) or after the change (time $t + 1$). The assertions about the kind of equilibria derived from COM are made supposing that the taken equilibrium action is A-type and engaging at time t (which in fact, given DR and RF implies that it will be A-type and engaging even at time $t + 1$.

Third, the fact that the DM may eventually reveal his final choice also means that, any inconsistencies as we consider them here are temporary. This suggests that the DM "solves" his inconsistencies by making choices. While it has been suggested that apparent inconsistencies that can be observed in people's behaviour in experimental settings can be resolved with repeated exposure to the market environment (e.g. Cherry et al. 2003), we find that all individuals need to do, if they cannot commit to a particular action, is to engage in a sequence of different actions and to change their behaviour. The best way of resolving conflict and inconsistencies is to continue choosing different actions, even inconsistent ones. John Stuart Mill (1859) suggested that exposure to choice helps to make better choices. Our theory shows that, in fact, exposure to choice helps to make more consistent choices.

We are of course aware that our theory will need to be challenged from at least two different perspectives. First of all, what happens if the DM does not have single-peaked motivations, but motivations with a different or more general structure? Second, we will have to consider decision problems that involve at least two if not more dimensions in which possible motivation conflicts can take place. This would mean that we also have to consider conflicts between choices along different dimensions. This may be the case when actions cannot only be considered "unidimensionally" as we suggest here. Consider the following example: Remember the person that evaluates actions on the dimension "healthy lifestyle". Suppose he has the peak of P at 0.4 and the peak of S at 0.8. Suppose this person has not been able to fulfill his self-image a lot: he prefers to eat healthy food that he cooks at home, but has not managed to practice sport regularly. His SQ lies somewhere at, say, 0.5. He has now been offered a job in which he could earn much more than what he currently earns, but which would imply that he could cook much less often at home and that he had to eat ready-made food in the cantine. The final action "choosing or not the job" can be split in two dimensions: "healthy lifestyle" and "earning money". Clearly, choosing the job would be a D-type action on the dimension "healthy lifestyle" because it would be worse in all aspects with regard to the current SQ, but an A-type action on the dimension "earning money". Whether or not the person will accept the job is a matter of further analysis. However, we do think that so far, simply with the assumption of two single-peaked orderings of motivations over one dimension, we have been able to provide new and valuable insights into choice problems of individuals.

Besides extending the number of dimensions, another natural line of development of our model lies on the assumptions about how motivations would reasonably change as a consequence of the individual's decisions. We have proposed two axioms, RF and DR, that we believe are psychologically well founded and fit well with numerous interpretations of conflicting motivations. The model certainly provides the framework to explore the consequences of assuming other assumptions about motivation change that in certain contexts might be considered more appropriate.

Another point worth studying is to consider in more detail the welfare implications of choices under conflicting, and changing, motivations. For example, how the DM comes to his final choice is different in various cases discussed above. In the case where the two peaks converge, this means that the pleasure motivation will become identical with the DM's self-image. While this may appear to be the perfect state with no more conflicts, it is important to note that this situation can arise because the agent also engaged in dissonance reduction (e.g. **CCM2**). While dissonance reduction may be a useful psychological tool to alleviate unpleasant internal states, in our case it does also mean that the S-ordering moves closer to a chosen action. Depending on the context in which such actions take place, this may have a more or less positive impact on the DM's welfare. In particular, such a situation may have a touch of what Jon Elster or Amartya Sen may call an "adaptive preference" and would mean in our context that one's personal goals are (for whatever reason) systematically downgraded.

Other welfare issues could be considered as well. But what this brief outline of futher research questions certainly suggests is that our framework allows to gain new insights into the underlying "forces" that lead people to make decisions and would thus provide a good tool to explore this welfare dimension in more depth.

7 References

Acland, Dan and Matthew Levy. 2010. "Habit Formation and Naiveté in Gym Attendance: Evidence from a Field Experiment", *Working Paper*, retrieved at: http://isites.harvard.edu/fs/docs/icb.topic734947.files/HabitNaivete.pdf

Ainslie, George. 1992. *Picoeconomics: the strategic interaction of successive motivational states within the person*, Cambridge: Cambridge University Press.

Akerlof, George and Rachel Kranton. 2000. "Economics and Identity", *Quarterly Journal of Economics*, 65(3), pp. 715-753.

Akerlof, George and Rachel Kranton. 2010. *Identity Economics*, Princeton: Princeton University Press.

Baigent, Nick. 1995. "Behind the veil of preferences", *The Japanese Economic Review*, 46(1), pp. 88-101.

Beattie, Jane and Sema Barlas. 2001. “Predicting Perceived Differences in Tradeoff Difficulty”, In Weber, Elke, Baron, Jonathan and Graham Loomes (eds.), *Conflict and Tradeoffs in Decision Making*, Cambridge: Cambridge University Press.

Bernheim, Douglas B. 1994. “A Theory of Conformity”, *Journal of Political Economy*, 102(5), pp. 841-877.

Black, Duncan. 1958. *The theory of committees and elections.* Cambridge University Press, Cambridge, MA.

Charness, Gary and Uri Gneezy. 2009. “Incentives to Exercise”, *Econometrica*, 77(3), pp. 909-931.

Charness, Gary and Matthew Rabin. 2002. “Understanding Social Preferences with Simple Tests”, *Quarterly Journal of Economics*, 117(3), pp. 817-869.

Cherry, Todd, Crocker, Thomas and Jason Shogren. 2003. “Rationality Spillovers”. *Journal of Environmental Economics and Management*, 45(1), pp. 63-84.

Davis, John. 2003. *The Theory of the Individual in Economics: Identity and Values*, London: Routledge.

Davis, John. 2011. *Individuals and Identity in Economics*, Cambridge: Cambridge University Press.

Deci, Edward. L. and Ryan, Richard. M. 1985. *Intrinsic motivation and self-determination in human behavior*, New York: Plenum.

Elster, Jon. 1989. *Nuts and Bolts*, Cambridge: Cambridge University Press.

Elster, Jon. 2000. *Ulysses Unbound*, Cambridge: Cambridge University Press.

Elster, Jon. 2007. *Explaining Social Behaviour: More Nuts and Bolts for the Social Sciences*, Cambridge: Cambridge University Press.

Falk, Armin and Urs Fischbacher. 2006. “A Theory of Reciprocity”, *Games and Economic Behavior*, 54, pp. 293-315.

Fehr, Ernst and Klaus Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation”, *Quarterly Journal of Economics*, 114(3), pp. 817-868.

Festinger, Leon. (1957) *Theory of Cognitive Dissonance.* Row, Peterson and Co.

Frey, Bruno. 1994. "How intrinsic motivation is crowded out and in", *Rationality and Society* 6(3), pp. 334-352.

Inada, Ken-Ichi. 1969. "The simple majority rule", *Econometrica*, 37, pp. 490-506.

Kalai, Gil, Ariel Rubinstein and Ran Spiegler. 2002. "Rationalizing choice functions by multiple rationales", *Econometrica*, 70 (6), pp. 2481.2488.

Kirman, Alan and Miriam Teschl. 2004. "On the emergence of economic identity", *Revue de Philosophie Economique*, 9, pp. 129-156.

Levi, Isaac. 1986. *Hard Choices: Decision making under unresolved conflict.* Cambridge: Cambridge University Press

Stuart Mill, J. 1859. "On Liberty", London: Parker.

Moulin, Herve, 1980, "On strategy-proofness and single peakedness", *Public Choice*, 35, 437-455.

Ryan, Richard and Edward Deci. 2000. "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions", *Contemporary Educational Psychology* 25, pp. 54-67.

Schelling, Thomas. 2006. *Strategies of Commitment*, Cambridge, Mass.: Harvard University Press.

Schwartz, Shalom. 1994. "Are There Universal Aspects in the Structure and Contents of Human Values?", *Journal of Social Issues* 50(4), pp. 19-45.

Selten, Reinhard. 1998. "Adaptation Aspiration Theory", *Journal of Mathematical Psychology*, 42, pp. 191-214.

Selten, Reinhard. 2001. "What is Bounded Rationality", In Gerd Gigerenzer and Reinhard Selten, *Bounded Rationality: The adaptive toolbox*, pp. 13-36, MIT Press.

Sen, Amartya. 1971. "Choice functions and revealed preference", *Review of Economic Studies*, 38, pp. 307-317.

Sen, Amartya. 1973. "Behavior and the Concept of Preference". *Economica* 40 (159), pp. 241-259.

Simon, Herbert A. 1955. "A behavioral model of rational choice", *Quarterly Journal of Economics*, 69, pp. 99-118.

Simon, Herbert A. 1956. “Rational choice and the structure of the environment”, *Psychological Review*, 63, pp. 129-138.

Steedman, Ian and Ulrich Krause. 1986. “Goethe’s *Faust*, Arrow’s Possiblity Theorem and the individual decision-taker”, in Jon Elster (ed.), *The multiple self*, pp. 197 -231, Cambridge: Cambridge University Press.

Stigler, George. J. and Becker, Gary. S. 1977. “De Gustibus Non Est Disputandum”. *The American Economic Review*, 67(2), pp. 76-90.

Sugden, Robert. 2004. “The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences”, *American Economic Review*, 94(4), pp. 1014-1033.

Tversky, Amos. 1972. “Elimination by aspects: A theory of choice”, *Psychological Review*, 79(4), pp. 281-299.

Tversky, Amos and Daniel Kahneman. 1991. “Loss Aversion in Riskless Choice: A Reference-Dependent Model”, *Quarterly Journal of Economics*, 106(4), pp. 1039-1061.