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

February 22, 2012

Work in Progress!

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 some motivation change. We show first that under the no motivations change case, certain choice behaviour that appears to be inconsistent from the standard rational choice point of view may be explained in our framework as the outcome of conflicting motivations. Afterwards, under the motivations change case, we present two psychologically-flavoured assumptions about how motivations are influenced by choices and also a weak condition of rationality. We show that, under such three conditions motivation change leads to less inconsistent choices and not to more inconsistencies as one may think. This means that it is only by choosing different actions that conflicts between two motivations can eventually be resolved and the definite and final preference for an action can be revealed.

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

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. 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 choice 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, she may not keep with this decision and drink more wine. 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 that any clear structure would be detectable. Some economic field experiments have tested what sort of incentives may

---

1A 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).
lead to binding decisions (or “habit formation” as they call it) and thus to a more regular and consistent (re)application of previous choices (Charness and Gneezy 2009; Acland and Levy 2010). 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 choice set (i.e. my new year’s resolution was to go running at least 3 times a week. Now it’s 7 o’clock in the morning of the day on which I should go running, but once again I have not made my mind up of whether I continue sleeping or whether I get up and ready for my workout...).

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). 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. 2 Aggregating motivations into a single ordering actually implies that there exists a systematic way to solve every conflict, and this is precisely what we do not want to presuppose. The purpose of this paper in particular 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).

Two important features of our approach are that it is purely ordinal and behaviourally axiomatic, in the sense that the axioms we propose have psychological foundations. This is to say that, on the one hand, we introduce certain axioms describing the behaviour of the individual based on some well-known psychological principles. On the other hand, though motivations here 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.

In a first part of our paper we consider choice under conflict with given and stable motivations. Later, we introduce the possibility of motivation change on the basis of two

---

2Thus we do not propose to attribute particular weights to the motivations, or to aggregate them axiomatically such as individual preferences are aggregated into a social preference ordering as in Steedman and Krause 1986.
psychological principles, namely “reinforcement” and “dissonance reduction”. In both cases motivations do 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 finding is that inconsistencies in a standard rationality sense are narrowly linked to the existence of conflict between motivations and should therefore not be considered as an anomaly. It also turns out that we can find a final and stable ATC-preference based on motivations but only under certain circumstances. More specifically, we do not need to presuppose the existence of an ATC-preference before the individual makes some choice, but it turns out that in some cases the individual may finally reveal an ATC-preference 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 engages in a sequence of actions. Section 5 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. The set of alternatives from which a person is going to choose one option is a set of actions. Hence, 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 where 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 ranking is at first always increasing and then always decreasing. This assumption is an important simplification, together with the assumption that the two motivations are single-peaked orderings over the same dimension. 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 be, in general, rationalizable by means of a single peaked preference or even a complete and transitive preference. 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.

The dimension over which both motivations are ordered could be, for example, the intensity with which a certain action is carried out. The peak of the $P$-ordering on the dimension “jogging” for a person who is not too fond of doing sport but who would like to be slim and fit, could then be, say, somewhere in the first third of the dimension (going jogging but not too fast sometimes on Sundays), whilst the $S$-peak could be situated somewhere in the second half of the dimension (get up early before going to work three times a week to do some jogging). This would clearly indicate a situation of “tension” as to what level of intensity is the best for that person. “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.

The dimension along which we consider actions could also be seen as something more complex than just 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 an hour and sleeping exactly 7 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 single-peaked fashion in terms of both, $P$ and $S$ according to that dimension.

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 $X$.

We will denote the class of single-peaked orderings that can be defined over $X$ by $\Sigma$. In our setting, we will work with two single-peaked orderings defined over $X$, pleasure, $P$, and the self-image, $S$, of a person, where 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 dimension for which $P$ and $S$ are single peaked will be normalized to the interval $[0,1]$ and actions will thus be formally represented by points in the interval $[0,1]$. The fact that $P \neq S$ means that there is potentially a kind of conflict to be resolved within the individual.

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

3Let $Q$ be an ordering (transitive and complete binary relation) defined over $X$. Let $b(Q)$ the best alternative of $X$ according to $Q$. Let $L$ be any linear ordering defined on $X$. We say that two alternatives, $x$ and $y$ are consecutive in $L$ if $xLy$ and there does not exist another alternative $z$ such that $xLzLy$. Then, we say that $Q$ is single-peaked if there exists a linear ordering $L$ such that, for any two consecutive alternatives $x$, $y$ in $X$, $b(Q)LxLy$ or $yLxLb(Q)$ implies $xQy$ (see, as basic technical references, Black 1958, Inada 1969, or Moulin 1980)
that is above the current level of pleasure (S-fulfilment) is pleasant (S-fulfilling), anything below it is unpleasant (S-non-fulfilling). Formally, since \( P \) and \( S \) are complete and transitive orderings defined over a connected subset of \( \mathbb{R} \), we can define a numerical function for \( P \) and \( S \), \( \ell_P \) and \( \ell_S \) respectively. It should be remembered that \( \ell_P \) and \( \ell_S \) are not utility functions because \( P \) and \( S \) are not preference relations. They should be interpreted as numerical measurements of the degree of pleasure and S-fulfilment 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 \( \ell_P \) and \( \ell_S \) is purely ordinal, that is, they are supposed to be unique up to any monotonic transformation.

The definitions above 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 assume that it does not provoke any conflict to the decision maker when confronted with the SQ.

**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. Any action chosen from within this set of actions thus creates a kind of conflict within the person as one of the motivations will not be satisfied.

**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 experience conflict 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.

The four kinds of actions are graphically represented in Figure 1. Without loss of generality, all figures are drawn such that the peak $\hat{P}$ lies to the left of the peak of $\hat{S}$.

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

---

4For 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.
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 type of actions plus the $SQ$ make a complete classification of 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.

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.

**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 this 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 $\ell_P(x) > \ell_P(SQ)$ 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, then it is not difficult to see that (e.g. imagine truncated $P$ or $S$ peaks such that according to the $SQ$ all actions are pleasant or $S$-fulfilling to the right of the $SQ$) there are no $B$-type actions if such a $z$ does not exist, and no $C$-type actions is such a $y$ does not exist, or there are neither $B$ or $C$-type actions. In any case the impossibility for both types of actions to co-exist would be proved. 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 $S$, for every $x \in (\hat{P}, y]$ $\ell_P(x) \geq \ell_P(SQ)$ and $\ell_S(x) > \ell_P(SQ)$, 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 $z(y)$.

We can also prove that if $SQ > \hat{S}$ then $B$ and $C$-type actions are incompatible following analogous steps. $\square$
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 $x < SQ$. But, now given the single-peakedness of $S$ and given that $SQ < \hat{S}$ we have that, for every $x < SQ$, $\ell_S(x) < \ell_S(SQ)$, 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, 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 enough strict diet for what she considers 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 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 agent either makes a more strict 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. 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 Figure 2(a). However, at a certain moment, for any actions that are even further to the right of the dimension the individual has to pay a price in terms of at least one of the two principles and $A$-options will not any longer be available. It can be the case that the action is not pleasant any more but it is worthy in terms of achieving 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). In any
case, by choosing an action further to the right of the dimension, $A$-type actions will not any longer be available, and we find either $C$-type actions as in Figure 2(a) (when actions become unpleasant but are still $S$-fulfilling), or $B$-type actions as in Figure 2(b) (when they become non-$S$-fulfilling but are still pleasant), or eventually $D$-type actions (as for example action $w$ in Figure 2(b), when these 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 prevent, in principle, that the agent takes a decision that puts into conflict the pleasure and the self-image (by choosing either $B$-type or $C$-type actions), but the kind of conflict she is facing is going to be of either of one or the other type. It is important to remark that, in our model, even if $A$-type actions are available, we take it that the DM does not necessarily have to choose $A$-type actions. We assume as perfectly possible that she chooses a $B$- or a $C$-type actions if she values highly enough the gain in the corresponding motivation by choosing such an action, even if such an action involve a loss in another motivation, and even if there exist other alternatives that involve, for example, small gains in both motivations.
3 Choice with Given 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 is compatible with a “satisficing” approach to decision making, 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, namely linking inconsistencies with the existence of conflict between motivations. More precisely, we assume that choices are the output of two conflicting motivations (represented by $P$ and $S$) and the status quo.

Let $C : 2^X \times X \leftrightarrow X$ be a choice function that assigns to a choice set and a status quo a unique choice. Thus $x = C(K, z)$ will be read as “$x$ is the choice from set $K$ when the status quo is $z$” for all $K \subseteq X$. We will assume that, if $K$ is the choice 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.

In order to make a first axiomatic approach to which actions should reasonably be excluded in our framework, we start by presenting the following simple axiom:

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

The axiom above just expresses that, when compared with the status-quo, an action $x$ is worse with respect to the two motivations under consideration, then it will never be chosen. The axiom is rather weak because its domain is rather small in two senses: on the one hand it only concerns comparisons of actions with the status quo, and on the other hand, it only applies when the alternative to the status quo is worse (not better) with respect to the two motivations. Given the classification of actions above, it is clear that SQLM directly leads to consider every $D$-type action as 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 \( x \). We start with a narrow definition of this kind of phenomenon in connection with the status quo, and 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 “regrets” not to be realising the other action. This dithering is the consequence of the changing SQ, and not a consequence of the impossibility to rank two identical options, such as the two haystacks from which Buridan’s ass has to choose. Rather, one haystack has golden looking hay, but is smaller than the other haystack whose hay has not such an appetising colour. 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. More worldly situations are those of smokers for example, who 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 can get motivated to go, say, jogging, 3 times a week, only to suddenly lose their motivation and to go jogging only on Sundays.

**Result 3.3.** *Assume that SQLM holds. Then for some \( x, y \in X \), for some \( K \subseteq X \) \( y = C(K, x) \) and \( x = C(K, y) \) (status quo dependent choice reversal holds) if and only if \( y \) 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.** \( y \) is not a \( D \)-type by SQLM. \( y \) 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), where \( K = X \). Hence, for preference reversal to arise, whatever set \( K \) is, the SQ has to become 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 preference 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 will prove, by means of two examples, that both, \( B \) or \( C \)-type actions may lead to preference reversal. 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 \( x = SQ \) and that \( x \), when chosen again, is a \( C \)-type (\( B \)-type) action can easily be proven by the very definition of \( B \)- and \( C \)-type actions.

This result suggests that, in our model, this kind of choice reversal is strongly related with conflictual choices, that is, with \( B \) or \( C \)-type actions. This choice reversal reveals a kind of “Anti-SQ-bias”: the option that one does not have is always better. Jon Elster has used the “The grass is always greener on the other side of the fence” syndrom to explain what he calls counter-adaptive preferences. For Elster, this 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 experience conflicting motivations which they are not able to trade-off in any consistent way. This situation implies that individuals tend to focalise 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 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 other 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 \( P(S) \).

The idea of our choice reversal is related with Tversky and Kahneman’s (1991) loss aversion, where they value gains and losses 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 correspondence but simply state that for a particular case of choice reversal, gains must be valued more losses.
A translation that rationalizes the choice function exists is Independence of Irrelevant Alternatives status quo dependent choice reversal it will be violated. Its violation in fact will include status quo dependent choice reversal in which an action that is worse than the status quo is rejected when the status quo is added to a set of alternatives containing the action. Obviously, the status quo dependent choice reversal is not the only kind of inconsistency in choice that one can observe. The consistency condition that ensures that a preference relation that rationalizes the choice function exists is Independence of Irrelevant Alternatives (See, for example, Kalai, Rubinstein and Spiegler (2002)). Next we propose a direct adaptation to our context of that condition, 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 consider the following condition (IIA): $\forall K, T \subseteq X$, $K \subseteq T$, $\forall x \in K, T$, if $x = C(T, \cdot)$ then $x = C(K, \cdot)$. We will say that IIA is violated when, for some $K, T \subseteq X$, $K \subseteq T$, there exist $x \in K, T$, $y \in K, T$ such that $x = C(T, \cdot)$ and $y = C(K, \cdot)$. In that case we will also 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 $x$ being $y$ available, but in another situation where there are not additional actions and still both actions are available, the DM chooses $y$. Clearly enough, given the importance of the status quo in our framework, such a violation is rather plausible due to the effect of changes in the status quo when the set shrinks. Our aim is, however, to prove that any violation that may happen will be always linked to the presence of conflictive choices.

Here, unlike with the status-quo dependent choice reversal, even $A$-type actions might lead to inconsistency. In this case inconsistency might happen, not in relation with the status quo, but in relation with a third action that, having been rejected when the $A$-type action $x$ was chosen can be, however, chosen when $x$ becomes the status quo. For example, consider Figure 2(a) again. To simplify, assume that $X$ is the choice 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, it is admissible that the DM chooses $y$ even though $x$ is also available and thus engages in revealed inconsistency.

Compared with status quo dependent choice reversal, considering the violation of IIA opens the door to a large range of inconsistent behaviours. Given that, it is reasonable to think about further axioms concerning 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 $x \neq C(K, \cdot)$.

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. In other words, 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 $K$ contains actions between the two peaks, DM constrains the range of admissible actions to them, which might be either $A$, $B$, or $C$-type actions. However, we will continue assuming that the initial status quo can be to the left of $\hat{P}$ or to the right of $\hat{S}$.
Result 3.6. Assume that DOM holds, then

- If a pair of actions \( (x, y) \) leads to the violation of IIA, then \( y \) is either a \( B \) or \( C \)-type action.

- If for some \( x, y \in X \), if \( y \) is either a \( B \) or \( C \)-type action, then it is possible that \( x = C(T, \cdot) \) and \( y = C(K, \cdot) \) with \( K \subseteq T \) (they lead to the violation of IIA).

Proof.

Assume that \( x = C(T, z) \) for some \( z = SQ \), 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, \( x \in I \). By Result 2.6, there are no \( A \)-type actions any more. By DOM, \( y \) can never be a \( D \)-type action either. Therefore, it can only be a \( B \)-type or \( C \)-type action. (ii) \( T \cap I = \emptyset \). Then, if \( y \in K, T, \ell_p(y) \leq \ell_p(x) \) and \( \ell_s(y) \leq \ell_s(x) \) with at least one strict inequality, then by DOM \( x \neq C(T, \cdot) \), thus it is impossible that it induces a violation of IIA. Furthermore, by DOM, \( y \) can never be a \( D \)-type action. Again, it can only be a \( B \)-type or \( C \)-type action.

Now we have to prove that being \( y \) either \( B \) or \( C \)-type actions, the violation of IIA is possible under DOM. It is not difficult (but tedious) to show examples where \( x \) is either an \( A \), \( B \) or \( C \)-type action and \( y \) is either \( B \) or \( C \)-type and where revealed inconsistency holds. Obviously \( x \) can never be a \( D \)-type action by DOM.

Result 3.6 connects inconsistencies more closely with conflict. That is, from the second choice onwards, any violation if IIA is only arising because of \( B \) or \( C \)-type actions. So even if the first action is non-conflictive and improving in both dimension with respect to the original \( SQ \), all of the following actions will induce conflicts and potentially revealed inconsistencies. The next axiom establishes an even stronger connection between conflict and inconsistency.

Axiom 3.7. Sticky Choice (SC): \( \forall K \subseteq X, \forall x, x', y \in X \) such that \( \ell_p(x') \geq \ell_p(x) \) and \( \ell_s(x') \geq \ell_s(x) \). If \( y \neq C_t(K, x) \) then \( y \neq C_{t+1}(K, x') \) for all \( t \in \mathbb{N} \).

Like DOM, this axiom is of a purely ordinal nature and imposes some kind of consistency in the sense that if the decision maker reveals that she is not willing to choose an action \( y \) being \( x \) the \( SQ \), she should not choose \( y \) the next time against a \( SQ \) that is even better in both dimensions than \( x \). The intuition behind axiom 3.7 is not far from that of classical \( \alpha \) condition in rational choice. This means that a strict improvement in both dimensions involves higher “standards” and thus reduces the set of “admissible” actions. Therefore
something that was not chosen with a lower standard (in terms of a lower \( SQ \)) will not be chosen when the standard has risen. The consequence of this axiom is that once the DM is taking an \( A \)-type action, she will not deviate from it and stick to this decision.

**Result 3.8.** Assume that \( \text{DOM} \) and \( \text{SC} \) hold, then

- If a pair \( x, y \in X \) leads to a violation of IIA, then \( x \) and \( y \) are either \( B \) or \( C \)-type actions.

- If for some \( x, y \in X \), \( x \) and \( y \) are either \( B \) or \( C \)-type actions then it is possible that \( x = C(T, \cdot) \) and \( y = C(K, \cdot) \) with \( K \subseteq T \) (they may lead to violation of IIA)

**Proof.**

Assume that \( x \) is an \( A \)-type action, then, it is better than the status quo in terms of both, \( P \) and \( S \). By \( \text{SC} \) this prevents any other alternative \( y \) that was available at the first choice to be chosen, and therefore makes the violation of IIA impossible. Moreover, by \( \text{DOM} \), \( x \) cannot be a \( D \)-type action either. 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

![Diagram](image-url)
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. In figure 4(a) we see that $x$ is chosen as 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 SC. 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 SC.

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 by imposing this condition status quo dependent choice reversal turns out to be only associated with $B$ and $C$-type actions, that is, with situations of conflict.

Results 3.6 and 3.8 come to a conclusion, which is similar to that of Result 3.3. We find a link between conflicting choice and inconsistency. Since the violation of IIA 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 ruling out $D$-type actions. If we impose DOM, apart from maybe the first action, all other actions are either $B$ or $C$-type actions. If in addition to DOM, we impose SC, we have even a stronger link between Result 3.8 and Result 3.3 in the sense that inconsistencies can only arise in the presence of $B$ or $C$-type actions.

In the following, we will turn to two explicit 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 and extent of Result 3.8 providing an interpretation of Ainslie’s party example.

**Example 3.9. 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 it would please him 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, is he going to act according to his decision and goes home (and thus 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 yet 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 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 “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 restricts the set of actions to which the person will be able to commit before he goes to the party. Assume that the person takes the decision to go to the party until 10pm (action x), and that this decision can be considered as an A-type action (which is possible given that the SQ is to the left of the peak of P), then according to our analysis above, when 10pm approaches, by SC the person will not change his mind and choose to stay longer (action y), but will leave the party without experiencing any kind of further conflict. He thus would not be revealing any kind of inconsistency. In fact, the decision to leave at 10pm consisted in such a big improvement with respect to the previous SQ, that the person refrains from choosing any other action that would involve necessarily some further conflict. This situation is represented in figure 5(a).

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. This person may well reconsider his previous decision to leave 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. However DOM restricts again the set of possible inconsistent actions the person could choose. The reason for this inconsistency is that the first choice the person has undertaken was already a conflicting one for him in the first place and the conflict is continuing to matter for the person once he is at the party and he has to implement his previous decision. Hence given the conflictual background of his decision, it may well be that he is going to decide against his previous interest of passing the exam. In fact, the decision to go to the party for a responsible person is already a B-type
action, hence a conflictual choice that goes against his self-image and this may induce further conflicts. This example does therefore suggest, even though it seems counterintuitive, that the person who is more responsible has more chances to act inconsistently than a person who is less responsible.

**Example 3.10. 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 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. The individual’s associated pleasure of putting in more effort and gaining more is at first increasing, but then decreasing. $P$ can therefore be
considered to represent the individual extrinsic motivation. $S$ on the other hand represents
the pleasure and satisfaction the person receives from doing the work itself, with an anal-
ogous interpretation. For example, it may be that the peak of $S$ will be more to the right
of the dimension, the more the job matches the person’s idea of what sort of professional
activity he wants to do. $S$ thus represents in this example the person’s intrinsic motivation.

Suppose now that the person’s $P$-peak lies to the left of the $S$-peak. This represents a
situation in which the pleasure the person receives from being paid for the particular job he
does is actually lower than the pleasure he receives from doing the job for the sake of the
job itself. Suppose also 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 pleasure from the wage increase than before.

We can differentiate between two situations. In the first situation, represented in fig-
ure 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 $\hat{S}$.  

Figure 6: Intrinsic and extrinsic motivation example
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 $B$ or $C$-type actions in addition to simply remaining at the $SQ$. Here it may be very likely that the person is actually going to decrease his effort. While he would lose some of his pleasure from doing the job in itself, he will be earning more money than before for a lower effort level.

### 4 Endogenous Motivation Change

What is now interesting to analyse are sequences of actions. 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 means that the status quo changes, 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.\footnote{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.} We express how we assume choices will affect motivations 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 that there may be other psychological influences on motivations too with different effects as the ones we propose. Moreover, in this section, we
will assume that every action will be available at any moment. Hence the opportunity set is $X$.

The first psychological axiom is called **reinforcement**. In this 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 previous status quo, the person experiences a form of dissonance that he wants to reduce. This is translated into a change of ordering $S$, in a way such that the individual is going to render it more coherent with his choice.

A short comment may be necessary here. It could be criticised that reinforcement and dissonance reduction are two psychological phenomena that can be associated with pleasure and self-image, but not necessarily with other motivations. Thus, the theory developed in this section depends on these motivations and cannot be applied to others. This is only partly correct. It may be true that reinforcement applies only to certain motivations, but not to pleasure only. It may be associated with the enjoyment of an activity where the person increases enjoyment by engaging in that activity and vice versa (creativity, rational thought, playing an instrument). This would be similar to the idea of habit formation (Stigler and Becker 1977). If a person is driven by courage or curiosity, it may also be that courage and curiosity increases with an accumulated history of actions done on the basis of those motivations. Dissonance reduction does also not need to be associated with self-image, but can be of various qualities and origins, such as altruism, social norm following, 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 curiosity on the one hand and obedience or honoring parents on the other, ambitiousness could conflict with a particular social order, etc. However, what is clear is that we propose a theory that assumes particular pairs of conflicting motivations where it makes sense to assume that reinforcement and dissonance reduction applies.

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, 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 $\hat{A}(Q,x)$ 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 qualitatively affected. Formally:

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

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

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. This adaptation is slightly related to the idea of “aspiration adaptation” (Selten, 1998) according to which, like in our case, taking actions produces a certain adaptation in aspirations. However, as will be seen soon, our idea of adaptation is more psychologically founded. 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.

Thus, we can present now the following two axioms:

**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 might 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 $\hat{P}$. Meanwhile, all the actions that
A individual will have to "rationalize" her self image less than the current status quo (see figure 8(b)). The idea is that the DR incorporates the findings of psychological research on cognitive dissonance following dissonance that was created by engaging in an action that distances her from her self image.

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 as in our context.

Jon Elster (1989, chapter 9) has given particular attention to this experience. For him, as in our approach, RF is not in general a conscious goal of action. However, for Elster, RF means that the fact that an action has pleasant consequences 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 there is no indication whether the person is therefore going to choose it more often or not, given that the person will still need to decide between two (competing) motivations.

The second motivation change is dissonance reduction.

**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 $\ell_S(x) < \ell_s(SQ)$.

In this case, $S$ will change if (and only 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" 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

---

6The 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 as in our context.
or more of them are in conflict with each other. By changing the $S$-ordering, the person accommodates 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 once more for any chosen action.

As a corollary, 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 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.

It has to be noted 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 it 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 $x \neq \hat{P}$, thus, by RF, if $y$ is taken by the agent $P$ 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, at the same time is self-image fulfilling. This is a situation of no-conflict as both motivations are fulfilled, with the person’s pleasure motivation even to its highest degree. In any other case however, engaging in actions will necessarily trigger motivation change.

All together, taking into account that $P$ can move either left, right or not move, and so can $S$, we have nine possible combined changes in motivations (CCMs). However, if we impose our two psychological axioms and also DOM, then we restrict the set of possible CCMs.

Result 4.5. Assume that RF and 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 no change either in $S$ or $P$.

Proof. As we said, there are nine possible CCMs. The three combinations where $P$ move 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 action the individual may take will contribute to reduce the distance between the two peaks. The two motivations are becoming closer, reducing the set of conflictual actions.

**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 at least two different actions, 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).

We can now make a global relationship among the implications of all the results obtained up to this point. Result 4.4 shows that, under DOM, acting successively leads necessarily to motivation change, and Result 4.5 shows that, again under DOM, such a motivation change is of a kind that reduces the distance between the two motivations. Furthermore, according to Result 4.6 we have that such a sequence of actions is going to be of a conflictive nature ($B$ or $C$-type actions except for the initial choice). Moreover, in the previous sections we have presented some results that show that inconsistencies in choice only arise when conflictual actions are undertaken. Consequently, as the two peaks approach, the range of conflictual choices shrinks. Though one could have thought that allowing motivations to change would lead to a larger range of inconsistencies, our model provides the interesting
(and unpredicted) conclusion that if we allow for motivations to change, then the potential range of inconsistencies is going to be reduced - and eventually even inconsistencies as such.

Indeed, we may wonder about the possibility of full convergence of the peaks of $P$ and $S$, 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. Nevertheless, even in terms of infinite sequences of actions it may not be necessarily the case that they lead to full convergence of the peaks. Actually, under DOM, we can distinguish two ways how the peaks of $P$ and $S$ may approach each other.

1. **Full convergence in the limit**: The (infinite) sequence of actions $\{x^n\}$ makes both peaks, $P$ and $S$ 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 affect motivations and the change in motivations induces new actions, with the final result being a convergence of both peaks and the progressive disappearance of conflict in the limit when both peaks coincide. As a matter of fact, 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 $P$ and $S$ 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 and motivation change. However, it could be the case that, in the limit, the peaks do not converge to the same point.

As a meaningful case of partial convergence in the limit we could have the situation that, from a certain moment onwards, the action taken is continuously the same, that is, at a certain moment the DM decides to stick to the SQ and not to change her decision any longer. Since choosing the SQ does not induce dissonant reduction, $S$ is not going to change any more, and for that reason the convergence in the limit will never be full. This can happen if the DM has found the final optimal choice when taking into account the

---

---

$^7$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.
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. However, there is nothing in the model that prevents the DM to choose, eventually, any action within the gap. In other words, there are no elements in the model to ensure that the resolution of the trade-off is "definite" or "final". This is a very special case of partial convergence but there are others. For example, choices can be continuously made in the gap between the peaks, but it could be perfectly the case that \(\lim_{n \to \infty} \{P'(x^n)\} < \lim_{n \to \infty} \{S'(x^n)\}\). This would be representing a situation where there is a limit in the agent’s motivations’ mutability, which at a certain point becomes infinitesimal. In the full convergence case, given that the peaks converge to a point \(x\), we have the trivial solution that \(\alpha l_p(x) + \beta l_s(x) > \alpha l_p(y) + \beta l_s(y)\) for any \(y\) different to \(x\) and for any value of the weights \(\alpha\) and \(\beta\).

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 does obviously not lead to any convergence. Nevertheless, there are some restricted relationships that can be proved. 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 \(\lim_{n \to \infty} \{P'(x^n)\} = \hat{S}\). 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. Many other examples about how the different CCMs lead to alternative types of convergence can be shown, with their corresponding psychological interpretation, but presenting all the possibilities would be too tedious.

Finally, there is a remarkable situation. Imagine that, besides DOM, SC also holds and that, at the beginning of the decision process there are \(A\)-type actions (that is, the initial \(SQ\) is not in between the two peaks). Then, SC implies that if the first action \(x\) is an \(A\)-type, it leads directly to a final resolution of the trade-off, that is, no action different to \(x\) will be chosen (see Results 4.10 and 4.11 for details later). The DM makes in this case one choice which is her final decision and then sticks to it. The \(P\)-peak will move towards the chosen action, and even might in the limit coincide with it if the action is taken repeateadly, thus making it to the most pleasant action available. However there will always remain a certain distance between the \(P\) and the \(S\) peak, which may be interpreted as if the DM gave systematically more weight to her pleasure than to her self-image fulfillment. As long
as A-type actions were available originally, which are typically no-conflictive actions, it can
be said that in such a case, the DM never experiences conflict because he decides to choose
an A-type action which then becomes a habit.

As a corollary, we can observe that by imposing only DOM we can have convergence of
the peaks with an A-type action, and when that happens: (i) The SQ is at first not in
between the peaks, and (ii) such an A-type action can only be one and the first action of
the sequence. After that, we only have B or C-type actions. If we further impose SC, such
an A-type action will involve necessarily the final resolution of the conflict straight away.
Thus, since processes of convergence (full or partial) involve infinite sequences of actions
and decisions, it can be said that any process of convergence is the result of B or C-type
actions. Only by engaging in conflictual actions, can the DM solve the conflict. There is
no way around experiencing conflict if the DM wants to get rid of it. Many people may
actually stick to the status quo because they are used to it and fear that any change would
foster the conflict. Our analysis shows that this is not the case. Change, as has been shown,
may eventually lead to the solution of the conflict.

Before concluding, we should consider one more issue, namely whether the above de-
scribed convergence processes can lead to an equilibrium. We define an equilibrium as a
situation where the decision maker will never change her current action.

**Definition 4.7.** We will say that \( x \in X \) is an equilibrium if \( x = C_i(X,x) \) for all \( i \in \mathbb{N} \).

The next results show that, under our model equilibrium is possible, and illustrate dif-
f erent ways how, according to the elements of the model, an equilibrium can be reached,
under the two scenarios under examination: no motivation change and endogenous motiva-
tion change.

**Result 4.8.** Assume that SQLM holds and that motivations do not change. Then we can
ensure that \( x \) is an equilibrium if and only if \( x = \hat{P} = \hat{S} \)

**Proof.** Clearly, if \( x = \hat{P} = \hat{S}, \forall z \in X, z \neq x, \ell_p(x) > \ell_p(z) \) and \( \ell_s(x) > \ell_s(z) \), thus, by
SQLM \( z \neq C(X,x) \) for all \( z \neq x \). Thus, \( x = C(X,x) \). Since motivations do not change the
previous argument applies every time the agent makes a choice, thus \( x = C_i(X,x) \) for all
\( i \in \mathbb{N} \).

On the other hand, we have to prove that we can ensure that \( x \) is an equilibrium only if
\( x = \hat{P} = \hat{S} \). For that, we will prove that, assuming that \( x = SQ \), if \( x \neq \hat{P} \) or \( x \neq \hat{S} \), then
there exist some \( z \in X, z \neq x, \) such that \( \neg(\ell_p(x) > \ell_p(z) \) and \( \ell_s(x) > \ell_s(z)) \), and therefore
SQLM does not rule out that \( z = C(X,x) \): If \( x \neq \hat{P} \), then either \( \ell_p(x + \epsilon) > \ell_p(x) \) or
\( \ell_p(x - \epsilon) > \ell_p(x) \) (depending on whether \( x < \hat{P} \) or \( x > \hat{P} \) respectively). Similarly, if \( x \neq \hat{S} \), then either \( \ell_s(x + \epsilon) > \ell_s(x) \) or \( \ell_s(x - \epsilon) > \ell_s(x) \).

\[ \] \( \square \)

**Result 4.9.** Assume that SQLM holds and that motivations may change according to axioms RF and DR. Then we can ensure that \( x \) is an equilibrium if and only if \( x = \hat{P} = \hat{S} \)

**Proof.**

Following the logic of the proof of Theorem 4.8, if \( x = \hat{P} = \hat{S} \), \( \forall z \in X, z \neq x, \ell_p(x) > \ell_p(z) \) and \( \ell_s(x) > \ell_s(z) \), thus, by SQLM \( z \neq C(X,x) \) for all \( z \neq x \). Thus, \( x = C(X,x) \). By RF and DR such a choice will not affect either \( P \) or \( S \), therefore the previous argument applies every time the agent makes a choice, thus \( x = C_i(X,x) \) for all \( i \in \mathbb{N} \)

On the other hand, we have to prove that, assuming motivation change, we can ensure that \( x \) is an equilibrium only if \( x = \hat{P} = \hat{S} \). In this case, the fact that motivation may change does not affect the argument used in the proof of Theorem 4.8.

\[ \square \]

**Result 4.10.** Assume that SC holds and that motivations do not change, and take \( x = C(X,z) \) for any \( x, z \in X \), then we can ensure that \( x \) becomes an equilibrium if and only if \( x \) is an \( A \)-type action.

**Proof.**

If \( x \) is an \( A \)-type action when \( z = SQ \), then \( \ell_p(x) > \ell_p(z) \) and \( \ell_s(x) > \ell_s(z) \), and SC directly implies that \( \forall y \in X, y \neq x, y \neq C_i(X,x) \) for all \( i \in \mathbb{N} \), thus \( x = C_i(X,x) \) for all \( i \in \mathbb{N} \).

(proof to be completed)

\[ \square \]

**Result 4.11.** (Equilibrium under SC and motivation change). To be stated and proved.

**Proof.**

\[ \square \]

**Result 4.12.** (Equilibrium under DOM and no motivation change). To be stated and proved.

**Proof.**

\[ \square \]
Result 4.13. (Equilibrium under DOM and motivation change). To be stated and proved

Proof. □

According to this definition, the elements of our model provide two ways to reach an equilibrium: First, any sequence of actions that leads to a full convergence in the limit gives an equilibrium. In this case, the two peaks finally coincide with a unique undominated action. When the equilibrium is of this type, it is SQLM what ensures that no other action is going to be undertaken. Second, as we discussed above, if SC holds and A-type actions are available and one of them, $x$, will be chosen, then $x$ is an equilibrium. Unlike the previous kind of equilibrium, in this case it is an axiomatic condition, namely SC, that ensures that $x$ is going to be the final choice. On the other hand, any partial convergence will not give rise to an equilibrium in our model. This is because as long as non-dominated actions with respect with the SQ remain available, we have no tools in the model guaranteeing that the individual will not move from her SQ.

5 Conclusion

Our preceding analysis of choice under (potentially) 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 underlying conflicting motivations. If we assume SQLM, which is certainly a minimal assumption of rationality, status quo dependent choice reversal only happens because of conflictual choices. If we incorporate some additional assumptions of rationality, namely DOM or SC, then not only status quo dependent choice reversal but also the violation of IIA are actually the outcome of pure conflicts that are incorporated in $B$- and $C$-type actions. Both actions satisfy only one motivation at the expense of the other motivation and because of this, the DM may be torn between two actions belonging each to one of the two types of actions in order to experience satisfaction interchangeably in at least one of the two motivations.

Second, despite the intuition that motivation change may lead to a greater range of possible inconsistencies, what we find is that, again, if we impose some reasonable conditions, motivation change leads eventually to 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.
Third, the fact that the DM may eventually reveal his final choice also means that, in such a case, any inconsistencies as we consider them here are temporary. This also 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 the individual needs to do is to engage in a sequence of different actions and to change his behaviour. The best way of resolving conflict and inconsistencies is to continue choosing different actions, even inconsistent ones. In a similar way that, according to Stuart Mill, exposure to choice helps to make better choices (look the references) we can say that, under our theory, 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 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 $D$-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 are able to provide new and valuable insights into choice problems of individuals.

There is yet another point worth studying and this is to look at the choices with motivation change from a welfare point of view. For example, how the DM comes to his final choice is different in various cases discussed above. In the case 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. 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 further 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.

6 References


