

Integrating Emotion and Rationality in Behavioral Models of Decision Making

Horacio Arló Costa
CMU

Abstract

Recent experimental evidence (Damasio 1994), (Bechara, Tranel, & Damasio 1997), suggests that emotions, rather than being sand in the machinery of rationality, are a condition of possibility of rational decision. The main goal of this paper is to present a model of the functional role of emotions in decision making. The offered model applies ideas first presented in (Simon 1967) to the problem of rational decision. The model is then used in order to interpret the experimental evidence and in order to suggest possible applications in knowledge representation.

It is proposed that three aspects of choice require the kind of 'interrupting control' hypothesized by Simon: (a) the specification of a feasible set, (b) the determination of mechanisms for picking, rather than choosing, in ties, and (c) the use and selection of heuristics. The application of normative theories of choice, which are partition-sensitive, like the one presented in (Savage 1972), require at least, the specification of the first two parameters. Emotions seem to play a fundamental role in setting those parameters, and therefore in applying normative theories of choice.

Introduction: Interpreting the experimental evidence

Antonio Damasio has provided clinical evidence of illuminating value for the understanding of the role of emotions in decision making. For example, he showed that patients with damaged frontal lobes become emotionally flat and lose their ability to make decisions, while retaining their cognitive powers. A typical example (Damasio 1994) is the following scheduling task. The researcher tries to fix the patient's next appointment and suggests two dates a few days apart in a month's time. The patient then starts arguing about the pros and the cons and the maybes of the alternative days. But the decision never comes. After half an hour, any normal person would have tossed a coin or done something - anything - to cut the process short. This does not happen nevertheless with prefrontal patients. When, after more than an hour of hesitation the researcher tells the patient that he should visit him on the second of the two dates, he responds "That's fine," as though there had never been any problem. This and other gambling experiments (Bechara, Tranel, &

Copyright © 2003, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

Damasio 1997) seems to indicate that the capacity of experimenting emotions and the capacity for sound decision-making decline together in patients with prefrontal damage.

Some of the previous experiments have been interpreted as suggesting that emotions provide a 'supplemental principle' that 'fills the gap' between reflex-like behavior and rational action (DeSousa 1997). De Sousa, for example, writes: 'For a variable but always limited time, an emotion limits the range of information that the organism will take into account, the inferences actually drawn for a potential infinity, and the set of live options from which it will choose ((DeSousa 1997), p. 195).' So, in the scheduling experiments mentioned above, normal subjects rather than adopting a 'rational' stance and examining all the available evidence, make a shortcut and proceed to ignore irrelevant information.

By the same token, it is known that people always use rules of thumb in order to make shortcuts in decision. Frijda argues that: 'emotion may well enhance utilization of these heuristics, in view of the desirability of rapid action or, more generally, the restriction of range of cue utilization ((Frijda 1986), p. 121).' Complex decision problems might require the use of these heuristics.

Experts in rational choice and decision making have reacted with skepticism towards this manner of interpreting the data. A good example is the recent work on emotions by Jon Elster. Here is what Elster has to say concerning the previous manner of interpreting the evidence:

What we may observe here, however, is not emotion doing what reason cannot do, but rather emotion doing what reason could also do, only differently. De Sousa and others who argue along similar lines consistently *present a strawman of rational choice theory*, according to which a rational agent would always take into account of all possible outcomes of all possible options. These authors all assume that rationality amounts to what I have called elsewhere *addiction to reason* (Elster 1989).¹

And later on, commenting on Damasio's interpretation of the gambling experiments:

Once again, going by one's gut feelings is not the only way to cut through the maze of a complex decision

¹(Elster 1999),p. 290).

problem. One can also, for instance, flip a coin. Damasio might counter that this procedure is inferior to going by gut feelings, which enable one not only to make swifter decisions but also better ones. But the coin-tossing heuristic is only the simplest of many rules of thumb that are used in complex decision-making problems. The best known is perhaps Herbert Simon's idea of *satisficing*, embodied in such sayings as 'never change a winning team' and 'if ain't broke, don't fix it' (Simon, 1954). Also medical diagnosis and prognoses can be efficiently done by mechanical point systems that relay on a small number of variables. In fact, such methods almost invariably tend to perform better than intuition based on 'gut feelings' (Dawes et al., 1989). *In opposing gut feelings to hyperrational cost-benefit calculation Damasio is simply setting up a strawman.*²

So, Elster complains that both Damasio and De Souza conflate rationality with what he sees as a pathological behavior, or *hyperrationality* (Elster 1989), p. 17). In general, ignoring the cost of decision making is an instance of what Elster calls hyperrationality. Ultimately Elster's argument is as follows: even when agents guided by their 'gut feelings' can outperform what he calls hyperrational agents, action motivated by gut feelings can be, in turn, outperformed by *bounded* methods of decision making, like, for example Simon's satisficing. To the extent that these bounded methods can be seen as encoding a viable notion of rationality, action motivated by gut feelings fails to be rational. So, Elster is skeptical towards the idea that emotions might promote rational behavior. Even if emotions might not be seen as sand in the machinery of rationality, they might be seen as 'quick and dirty' mechanisms which guarantee prompt action, even when they might do so by recommending sub-optimal strategies. Even when the argument seems *prima facie* correct, it seems to me that it fails to give a comprehensive picture of the interplay of rationality and emotions, and to do justice to the theories recently offered by neuro-scientists and philosophers. One aspect in which the views proposed by De Souza and Damasio seem correct is in their insight as to what is the functional role of emotions in decision making. In order to see the point it might be useful to restrict attention to the scheduling experiment with prefrontal patients. Presumably this is not a case where complex calculations are needed. Therefore the alternative use of bounded methods of decision does not seem necessary – 'bounded' methods are usually proposed in order to deal with the intricacies of complex decision cases. In the following sections I shall focus on this kind of problems. My goal is to propose a theory of the functional role of emotions in decision making, which extends a view first proposed by Herb Simon in (Simon 1967). The theory should be able to explain what is going on in simple scheduling problems. A brief, but important detour touching foundational issues in decision making is needed in order to introduce the main ideas of the proposal.

²(Elster 1999), p. 295, italics are mine.

The structure of a decision problem

There are two important aspects of decision making, which are not usually covered by standard axiomatic accounts of decision (like the one offered in (Savage 1972)). One of them is the selection of a feasible set of events, another is the secondary criteria useful in order to break ties. It would be nice to clarify the first issue with the help of the authoritative presentation of Leonard J. Savage. In section 2.5 of (Savage 1972) Savage makes a series of interesting remarks concerning the nature of *consequences*, *states* and *acts* in a decision problem (acts are sometimes called 'options' by Damasio and other scholars).

To say that a decision is to be made is to say that one of two or more acts is to be chosen, or decided on. In deciding on an act, account must be taken of the possible states of the world, and also of the consequences implicit in each act for each possible state of the world. A consequence is anything that might happen to the person. Consider an example. Your wife has just broken five good eggs into a bowl when you come in and volunteer to finish making the omelet. A six egg, which for some reason must either be used for the omelet or wasted altogether, lies unbroken in the bowl. You must decide what to do with this unbroken egg ((Savage 1972), p.13)

Savage suggests that in this case it is perhaps not too great an oversimplification to say that the decision-maker must decide among three acts which one can call: A1: 'Break into bowl', A2: 'Break into saucer', and A3: 'Throw away'. The states of the world are partitioned only by considering whether the egg in question is or not rotten. In other words, we have S1: 'Good' and S2: 'Rotten'. When S1 holds true, we have three possible consequences, which we can label as follows: O11: 'Six-egg omelet', O12: 'Six-egg omelet, and a saucer to wash', and O13: 'Five-egg omelet and one good egg destroyed'. By the same token, if S2 holds we have O21: 'No omelet, and five good eggs destroyed', O22: 'Five-egg omelet, and a saucer to wash' and O23: 'Five-egg omelet.'

Now, notice that the set up of the decision problem might be inadequate. For example, the decision-maker might be unsure as to whether one rotten egg will spoil a six-egg omelet. It is easy to see that this might lead to refinements of the *partition of events* first proposed into three states rather than two. In fact, the cell S2 ('Rotten') is then refined into: S2a ('Rotten and omelet spoiled') and S2b ('Rotten and omelet not spoiled'). Savage says that 'it seems to me obvious that this solution works in the greatest generality, though a thoroughgoing analysis might not be trivial ((Savage 1972), p.15).' So, in principle partitions are always refinable until one hits the finest partition possible, conformed by 'descriptions of the world leaving no relevant aspect undescribed.'

In addition 'the formal description might seem inadequate in that it does not provide explicitly that one decision might lead to another ((Savage 1972), p.15).' So, it is easy to conceive a situation where the act 'break into bowl' is substituted by several, 'such as: 'break into the bowl, and in case of disaster have toast', or 'break into the bowl and in case of

disaster take family to a neighboring restaurant for breakfast ((Savage 1972), p.15).’ In other words, one might propose to treat the decision problem in such a way that the choice of a policy or plan is regarded as a single decision. In a much-quoted passage Savage puts this in a suggestive manner:

The point of view under discussion may be symbolized by the proverb, ‘Look before you leap,’ and the one to which it is opposed by the proverb ‘You can cross the bridge when you come to it.’ When two proverbs conflict in this way, it is proverbially true that there is some truth in both of them, but rarely, if ever, can their common truth be captured by a single pat proverb. One must indeed look before he leaps, *in so far as looking is not unreasonably time-consuming and otherwise expensive*; but there are innumerable bridges one cannot afford to cross, unless one happens to come to them.

Carried to its logical extreme, the ‘Look before you leap’ principle demands that one envisage every conceivable policy for the government of his whole life in its most minute details, in the light of a vast number of unknown states of the world, and decide here and now on one policy... Though the ‘Look before you leap’ principle is preposterous if carried to extremes, I would none the less argue that it is the proper subject of our discussion, because to cross one’s bridges when one comes to them means to attack relatively simple problems of decision by artificially confining attention to so small world that the ‘Look before you leap’ principle can be applied there. I am unable to formulate criteria for selecting these small worlds and indeed believe that their selection may be a matter of judgement and experience about which it is impossible to enunciate complete and sharply defined general principles ... On the other hand, it is an operation in which we all necessarily have much experience, and one in which there is in practice considerable agreement ((Savage 1972), p.16).

Savage elaborates more on small worlds in section 5.2 of (Savage 1972). As Savage explains, an unmitigated use of the maxim ‘Look before you leap’ posits a task ‘not even remotely resembled by human possibility ((Savage 1972), p.16).’ On the other hand, he proposes a mitigated version of the principle by appealing to what he calls ‘small worlds.’ Selecting those worlds is tantamount to select an adequate partition of events for a decision problem. It is quite clear that Savage considers that this selection process is presupposed in formulating a decision problem. In other words, the principles of rational decision apply *after* this selection is made. Setting up a decision problem requires in general specifying the acts (options), the relevant partition of events and the consequences. For the purposes of this analysis I am assuming that each of these parameters can be specified independently of the other two (some decision theories, like Savage’s for example, assumes that acts just are functions from states to consequences, so only two of the three parameters really matter). While the literature exploring the role of emotions in decision has taken into account how emotions influence the option set (the acts) and the consequences, little is said about the way in which emotions might control the

formation and eventual refinement of the partition of events. I’ll argue below that this aspect of a decision problem is crucial in order to understand scheduling problems.

Interruption and Emotional Behavior

Both Simon, in (Simon 1967) and Hebb, in (Hebb 1949) produced perhaps the first models of the role of emotions in cognition. These models, and the more recent literature mentioned above, tend to defend the idea that an emotion stimulus should be regarded as *interrupting*, rather than as *disrupting* behavior. In general the models in question see emotional responses as adaptive responses, which can help to optimally reach a goal or that can help to reach a goal ‘well enough’ (satisficing).

The central idea is that complex cognitive entities are not only multi goal systems but that they are also equipped with an *interrupt system* capable of prioritizing goals in real-time responses. So, an interrupting stimulus ‘has a whole range of effects, including (a) interruption of ongoing goals in the CNS and substitution of new goals, producing, inter alia, emotional behavior, (b) arousal of the autonomic nervous system, (c) production of feelings of emotion ((Simon 1967), p. 36).’ Simon then stresses that the corresponding responses are ‘largely adaptive, either because they are genetically determined or because adaptation has been learned ((Simon 1967), p. 36).’³ Notice that in this model subjects undergo emotional responses because of the cognitive action of the interrupt system.

Simon, nevertheless, did not study in particular the role of emotions in choice situations (he focused on cognition in general). But it seems perfectly possible to apply his general model to choice by stating (the empirical) hypothesis that in setting up and solving a choice problem there are three main features of choice that require interrupting control. The first is the selection of a feasible *partition of events* and consequences. The second is the determination of mechanisms for *picking*, rather than choosing in ties.⁴ When preferences are completely symmetrical, i.e. when one is strictly indifferent with regard to the alternatives, one can refer to the act of taking (doing) one of them as an act of *picking*, rather than choosing. Choosing is always choosing for a reason, and this presupposes asymmetry in preference.⁵ But it is quite obvious that there are ‘selection situations’ without preference. We follow established terminology in rational choice by calling those situations ‘picking situations’. It is clear, for example, that a normal subject in a scheduling situation ‘picks’ one of the possible dates, when completely indifferent between the possible meeting dates. I am proposing here that one of the functional roles of emotions in choice is to make a ‘rational selection’ possible in situations without

³Simon provides some caveats: ‘When emotion-producing stimuli are persistent as well as intense, they sometimes become disruptive and produce non-adaptive behavior.’

⁴A clear distinction between picking and choosing is made for the first time in (Ullman-Margalit & Morgenbesser 1972).

⁵Leibniz’s dictum is applicable here: ‘In things which are absolutely indifferent there can be no choice ... since choice must have some reason or principle.’

asymmetrical preferences. Finally the last element of choice that requires interruption control is the use of heuristics in order to solve complex decision problems for which straight calculation is sufficiently costly (and for which the feasible options and the partition of events are already determined). In this article I shall focus on the first two aspects of choice. Even when they seem important enough, there is little investigation of them in the current literature.

One can hypothesize that the first of the aforementioned features has not only logical but also temporal priority in deliberation. Having a space of feasible events and consequences (or, as Savage wants, having a set up in terms of 'small worlds') seems to be a condition of possibility of solving a decision problem at all. If this is the right analysis of the role of emotions in deliberation, various testable hypotheses can be immediately formulated. For example, concerning scheduling experiments, one can conjecture that frontal lobe patients whose abilities for solving small optimization problems are intact, will be able to choose without problems in decision problems clearly presented to them, where the partition of events and consequences are fully specified beforehand (and where choice is clearly determined by preference). Of course, an array of experiments can be postulated following the model just proposed (see (Arlo-Costa & Gasper 2002) and (Arlo-Costa 2002)).

Partition sensitivity

A decision theory might or not be *partition sensitive*. The basic idea is that a decision theory is partition insensitive if the expected value of an option is a function of the values of the various ways in which it might be true, quite independently of the way in which these ways are sliced up (i.e. quite independently of the way in which the corresponding partition of events might be defined for the option). Some normative decision theories are partition insensitive (like for example Jeffrey's (Jeffrey 1983)). Other theories, like Savage's are partition sensitive.

The issue of selecting a maximally specific description of the world relevant to the purposes of deliberation has been considered in some decision theories, which are partition sensitive. For example, in the theory considered in (Levi 1999) a partition of the space of consequences is called *irrelevant* if it assigns the same values to the refined consequences. Therefore, a set of possible consequences O is called *basic* as long as every possible refinement O' of O is an irrelevant refinement (relative to the value structures of O and O'). Selecting a basic partition for a decision problem is seen in this theory as a requirement of rationality. One can see it as a process of adjustment where various partitions are tried until, given a basic partition, further partitioning is considered irrelevant.

When it comes to determine the aspects of decision making sensitive to interruption control, and therefore the aspects of decision making where there is a concrete interaction between emotions and rationality, the process of selection of basic partitions seem one of the natural candidates to take into account. Salient partitions might be immediately selected, relative to a library of problem types, or the process of convergence to basic partition selection might be

cut short. Of course, there is no guarantee of optimality for the selection in question. But we can agree with Simon that selections tend to be adaptive. The patients in scheduling problems can be described as performing a sequence of irrelevant partitions of the consequence space (in the technical sense used by Levi).

In a word, there are various (normative) types of decision theories and the problem of partition selection is quite different in each case depending on whether the theories are partition sensitive or not. Perhaps the reader is persuaded at these point that partition insensitive theories lack a problem that one expects a decision theory to have ((Sobel 1994), p.161). So, I shall use here partition sensitive theories, like Savage's.

One can hypothesize a dual role of emotions (both with adaptive consequences). On the one hand, one can see emotions as arising as the result of interrupting stimuli terminating a search procedure (for a feasible partition) which is too costly (interrupting stimuli in scheduling experiments can be provided by doctors reminding the subject of the time already employed in solving the task). On the other hand, one can also hypothesize a more rapid mechanism where types of decision problems are matched with a salient feasible set in a more or less automatic manner. For example, scheduling problems can be associated by default with the obvious kinds of partition types. What is interrupted in this case is the very process of selecting an initial partition of events. A *salient* type of partition is immediately selected for a type of decision problem. The central functional role of the interrupt system, and, inter alia, of emotions, in this latter case is to produce and utilize the saliency of certain partitions of events (for certain types of decision problems).

The proposed model differs from the one arising from Elster's analysis in various crucial aspects. If the decision task is sufficiently complex we can agree with Elster that 'gut feelings can in principle be out-performed by bounded models of choice (even when they do better than hyperrational responses).' Nevertheless, the proposed model inherits from Simon's analysis the *prima facie* assumption that the emotional responses will tend to be adaptive.

On the other hand, if the complexity of the decision problem is low (like in a scheduling task) the proposed model predicts that the gut feelings associated with cognitive interruptions are an external sign of a process which contributes to select a salient partition of events and to solve eventual ties; two aspects of choosing that ideal theories of rational choice treat differently and that tend to be relegated as almost arational aspects of rational choice. Again the prediction in this case is that the functional role of emotions will be adaptive rather than disruptive. It seems clear that the scheduling experiments (when run on normal subjects) belong to this second type of decision problems.

So, one can perhaps conclude that concerning the central issue of the role of emotions in cognition, there is a position which diverges both from Elster's and to some extent, from the views defended by Damasio and DeSouza, and which, is also more optimistic than Elster's regarding the impact of

emotions in overall rationality.⁶

On the one hand, the position has some of its roots in the writings of Herbert Simon. On the other hand, I extended it here in a concrete application to choice behavior.

Feasibility and ties

The previous analysis focused on foundational issues related to two aspects of interrupting control: the selection of a feasible set of options and the solution of ties. As a matter of fact I have focused on feasibility in detriment of the analysis of ties in choice behavior. I shall say here something more about ties. Ullmann-Margalit and Morgenbesser have written some of the most illuminating passages concerning the distinction between picking and choosing:

When we are in a genuine picking situation we are in a sense transformed into a chance device that functions at random and effects arbitrary selections.

Often enough, or perhaps typically, what occurs in a selection situation you identify as a picking one is that you haphazardly focus your attention on some one of the available alternatives. Once you do that, however, then - by hypothesis - none of the other alternatives attracts you more, and there are no room for qualms or second thoughts.

Little is known, nevertheless, about the mechanisms that make possible to focus attention in situations of this type. An hypothesis that seems natural is that the focusing in question needs to be under interrupting control as well. A prefrontal patient facing the scheduling situation described above can be described as trying to transform a picking situation into a choosing situation by refining the current partition (if he is originally indifferent regarding the two possible days for the next appointment). But if he happen to be situated in a basic partition already, all refinements will be irrelevant, and potentially endless. Normal subjects can be seen, in contrast, as capable of recognizing interrupting stimuli (questions from the doctor, reminding you of the time already spent in deliberation, constitute an example) which terminate either in picking or in a sufficient amount of refining before the adoption by default of the current partition as basic (which, in turn, can lead to picking or to choosing).

⁶I offer a critical appraisal of Damasio's theory and of some of its critics in (Arlo-Costa 2002). It is enough here to mention in passing that our account does not seem incompatible with Damasio's theory. It is true, nevertheless, that Damasio's theory has focused almost exclusively on the role of emotions in trimming the set of options (or acts). It is clear in the case of scheduling experiments that this is not the problem (there are only two acts: making an appointment for one of the dates or for the other). The problem is how emotions modulate the partition of events used to solve the problem. Damasio is right in pointing out (see (Damasio 1994), p. 51) that the prefrontal agent seems unable to attribute values to acts. It is nevertheless unclear in his account why this is so. My hypothesis here is that even when the options are identified by the agent, the corresponding partition of events is constantly refined in order to find a *reason* to untie a tie. A more elaborate account of why this is the case is offered below when I analyze the role of ties.

In this paper I have chosen to present the relations between emotions and rationality in the context of a numerical presentation of decision theory. But much of what I proposed can be applied as well to ordinal (or qualitative) theories of decision making, which have been recently proposed both by philosophers and by researchers in computer science.⁷

References

- Arlo-Costa, H., and Gasper, J. 2002. Emotional effects on feasibility and heuristics in choice behavior: experimental report. Technical report, Carnegie Mellon University.
- Arlo-Costa, H. 2002. Motivation and feasibility: The functional role of emotions in deliberation. Technical report, Carnegie Mellon University.
- Bechara, A.; Tranel, D.; and Damasio, A. 1997. Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–1295.
- Damasio, A. 1994. *Descartes' Error*. New York: Putnam.
- DeSousa, R. 1997. *The Rationality of Emotion*. Cambridge, Mass.: The MIT Press.
- Elster, J. 1989. *Solomonic Judgements: Studies in the Limitations of Rationality*. New York: Cambridge University Press.
- Elster, J. 1999. *Alchemies of the Mind: Rationality and the Emotions*. New York: Cambridge University Press.
- Frijda, N. 1986. *The Emotions*. New York: Cambridge University Press.
- Hebb, D. 1949. *The Organization of Behavior*. New York: Willey.
- Jeffrey, R. 1983. *The Logic of Decision, 2nd edition*. Chicago: The University of Chicago Press.
- Levi, I. 1999. Value commitments, value conflict, and the separability of belief and value. *Philosophy of Science* 66:509–533.
- Savage, L. 1972. *The Foundations of Statistics, second revised edition*. New York: Dover Publications, Inc.
- Simon, H. 1967. Motivational and emotional controls of cognition. *Psychological Review* 74:29–39.
- Sobel, J. 1994. *Taking Chances: Essays on Rational Choice*. New York: Cambridge University Press.
- Ullman-Margalit, E., and Morgenbesser, S. 1972. Picking and choosing. *Social Research* 757–785.

⁷The issue of how emotions influence the use of heuristics in choice is, of course, quite important, but, given space limitations, I did not tackle this issue here. I hope to have persuaded the reader, nevertheless, that even if we circumvent this problem, much can be learned about the functional role of emotions by focusing on their impact in setting the space of feasible events and in making the act of picking possible in cases of symmetric preference.