Proceedings of the 1999 IEEE
International Symposium on Intelligent Control/Intelligent Systems and Semiotics
Cambridge, MA • September 15-17, 1999

Emotions and Intelligent Choice: Feasibility, Seriousness, and Admissibility

Michael A. Goodrich
3368 TMCB
Computer Science Department
Brigham Young University
Provo, UT 84604
mike@cs.byu.edu

Abstract

Emotions and rationality have often been treated as opponents, with emotions taking on the role of villain and rationality taking on the role of heroine. Recently, however, the positive role of emotions has received considerable attention, and resulted in the elevation of emotion to rationality's equal. With emotion's change of status, artificial and computational intelligence researchers have begun to model and use emotions as a constructive portion of designing machines and describing humans. Building on the idea that emotions are generated by past actions (or, at least by the consequences produced by past actions) and are used to determine future actions, it is desirable to create a computational framework capable of representing these dynamics. Employing the hypothesis that some aspects of emotional state result from positive influences and other aspects result from negative influences, we can determine conditions when an action is a feasible possibility, a serious possibility, and an admissible choice.

1 Introduction

Traditional approaches to decision making have relegated emotions to second class status or, worse, ascribed them to an interference in intelligent choice. The divorce of emotions from rationality is an artifact that finds its justification not in mathematical decision theory with its prominent use of emotion-like utility, but instead in a perception based on anecdotal evidence that emotion "muddles ones thinking". However, work by neuro-physiologists (in particular, Damasio [2]) and by researchers in affective computing (such as Picard) suggests that emotion is essential for intelligent choice. In reality, emotions rightfully pervade almost all intelligent decision making.

The isolation of rationality from emotion begs certain questions that must be addressed if engineers, scientists, and philosophers are to make the transition from volitionless automata to intelligent agents. For example, the perpetuation of the rationality-emotion dichotomy allows an agent to maximize expected utility, but simultaneously requires that preference patterns, utilities, and options somehow exist beyond the rational analysis of the agent [8]. Given the goal of reuniting emotion and rational choice in a theory of intelligent choice, we entertain two hypotheses regarding the role of emotion in rational choice:

1. Emotions are both local instantiations of the expected fulfillment of global goals as well as evaluations of the consequences from past actions.
2. Opposing motivations determine the set of options admissible for rational choice.

A computational framework for describing and, in future work, testing these hypotheses can be obtained by organizing situated rational agency (i.e., intelligent choice) into a multi-resolutional society of Minskian choice agents [9]. Using this multi-resolutional society, the determination (itself a choice problem) of utilities, options, and beliefs used by a choice agent can be facilitated by external (meta-choice) agents to produce results consistent with standards for situated rational agency.

2 Decisions in a Vacuum

Consider the standard formulation of a decision problem diagrammed in Figure 1. The elements of this decision problem include states of nature Θ that represent the external or internal states that affect the consequences of a decision, the set of observations X of the state of nature, the set of options U that are available to the agent, the set of consequences C that result when the agent chooses $u \in U$ when

0-7803-5665-9/99/$10.00 © 1999 IEEE

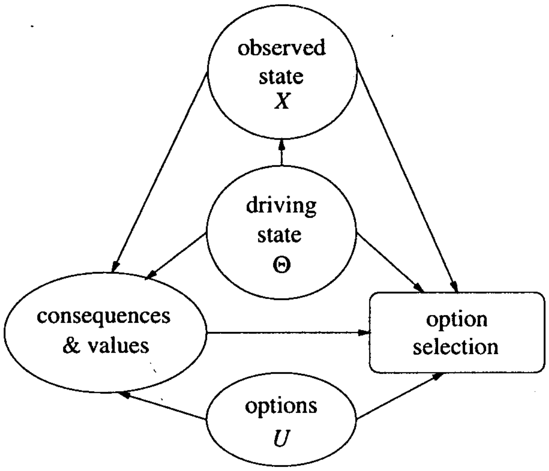

Figure 1: The choice problem. Arrows indicate either perceived (by the decision maker) or actual influence. For example, a consequence is the result of taking action $u \in U$ when $\theta \in \Theta$ obtains. Not all decision methodologies explicitly account for all influences. Typically, complex decision mechanisms seek to make all elements and influences explicit, whereas simple inference heuristics discard some of the influences or omit some of the elements.

the state is $\theta \in \Theta$ as described by a causative model[1] (i.e., consequence operator) $f : \Theta \times U \mapsto C$, preferences over consequences that represent encoded values, and a decision mechanism to select an option consistent with preferences. The traditional prescription of rational choice is that there exists a numerical utility function that represents preferred consequences $J : C \mapsto \Re$. The agent then chooses an optimal option u^* that maximizes the expected utility of the options

$$\begin{aligned} u^* &= \arg\max_{u \in U} J(c) \\ &= \arg\max_{u \in U} J(f(u,\theta)) \\ &= \arg\max_{u \in U} J'(u,\theta) \end{aligned}$$

where $J' = J \circ f$ is the composition of the consequence operator and the evaluation operator.

This description of rational choice is attractive because it produces a unique (up to an equivalence class) best decision u^* given an estimate of the state of nature and a preference pattern J or J'. However, this formulation does not address the following questions:

1. How does the agent determine a subset of U admissible for rational choice?
2. How does the agent determine the set of relevant states Θ?
3. How does the agent determine their value structure J?
4. How does the agent determine the mapping f between actions and consequences?
5. What is the tradeoff between information encoded in f and information encoded in J'?

If these things are known, then the rational choice problem can be mathematically formulated as a choice ("G" for "Get") operator

$$u^* = G(J, U, \Theta, f), \tag{1}$$

but, in the absence of these things no rational choice can be made. Thus, determining J, U, Θ, and f is essential; decisions in a vacuum are no decisions at all. In the following sections, the nature of J and the determination of a set of admissible options are considered.

3 Emotions: Local Instantiations of Global Objectives

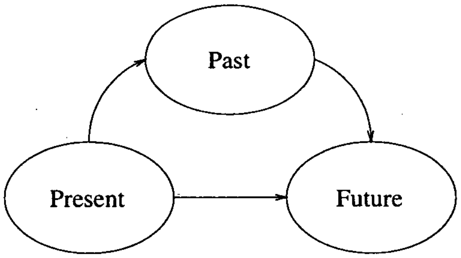

Figure 2: Interface between past, present, and future. Past experience, through explicit causative models or through utility elicitation, allows a decision-maker to map present observations and options into desirable future consequences.

As I interpret the philosopher Charles Peirce, meaning and therefore intelligence can only be present in a semiotic triad consisting of some kind of observation (firstness), some kind of consequent (secondness), and some kind of mapping from observation to consequent (thirdness) that turns firstness into secondness. Goal-directed agents capable of continued existence in real environments should have the capacity to respond to, interpret, and evaluate observations in terms of their capacities and skills. The key to doing this

---

[1] In the more general case, U can be a sequence of actions and f can be a mapping from this sequence of actions into a set of consequences.

is to allow lessons learned from the past (in the form of values, models of causal behavior, etc.) to turn observations from the present into acceptable future consequences. As diagrammed in Figure 2, the past (thirdness) facilitates the transformation of the present (firstness) into the future (secondness).

Thus, a minimum requirement for intelligence is a relationship between past[2], present, and future. To facilitate this relationship, there must be three phases for any choice problem: anticipating consequences, the "moment of truth" when choice is made, and evaluating consequences (see Figure 3). Anticipating future and evaluating past consequences are necessary stages in rational self-policing. By evaluating past consequences, a decision-maker is evaluating its past choices and is thus performing a third person (meta) evaluation of a "past self." If performance is inadequate or if superior alternatives are manifest then the decision maker should adapt its future behavior. By anticipating future consequences, a decision maker is evaluating its future states and is thus performing a third person (meta) evaluation of a "future self." If expected performance is inadequate or if superior alternatives are recognized then the decision maker should act accordingly.

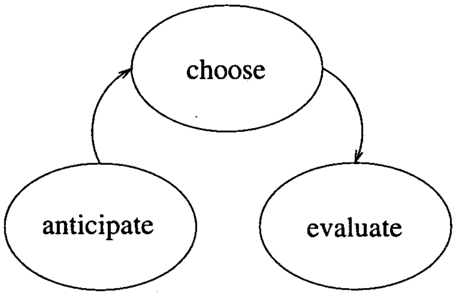

Figure 3: Three phases of the situated decision making. Choice is the *moment of truth* for a choice agent, and anticipation and evaluation are elements of (meta) rational self-policing.

The lesson we learn from this triad of situated agency is that much of reasoning is done in terms of either past experiences or expected future experiences. This can be extremely complex unless effective coping strategies are developed and used. A remarkably efficient coping strategy is to organize intelligence into modules appropriate for commonly encountered circumstances. We call these modules cognitive or behavioral skills and note that these skills determine the behavior of a situated decision maker. Such a decision maker can reason about the world in terms of the consequences afforded by these skills. With the emergence of multiple skills including the capacity for general-purpose problem solving, a decision maker can be capable of very sophisticated behaviors.

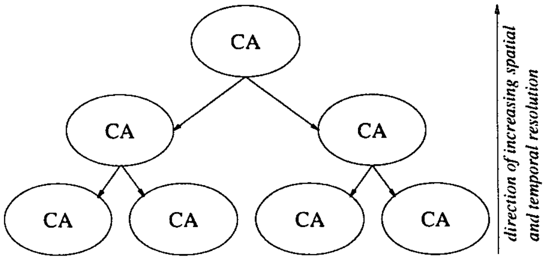

Figure 4: Multi-resolutional society of choice agents (CA's).

There is evidence suggesting that emotions depend on a hierarchy of motivational factors [6], and work suggesting that intelligence can be organized into multi-resolutional societies of Minskian agents [3]. With this as a background, consider the multi-resolutional organization of intelligence depicted in Figure 4 in which multiple choice agents (CA's) are organized into a hierarchy with ever increasing scope (for example, choice agents at the top of the hierarchy operate with longer temporal planning/evaluation horizons than choice agents at the bottom of the hierarchy). One result of the multi-resolutional society of choice agents is that the value operator J depends on the objectives of its parent choice agent. This value function is produced by the parent agent and is used by the child agent to select a rational choice. Thus, to the parent the value function is Peircean secondness whereas to the child the value function is Peircean firstness. Within the child agent this value function can be used to produce action, but it is an impression (firstness) rather than the result of the child agent's interpretation (secondness). In this way, the value function is an emotion (a firstness)[3] to the child agent that can be used to generate a decision but cannot be interpreted without consulting the parent agent; the value function is an emotion and this emotion is a local instantiation of the parent's objectives in the context afforded by the environment.

Decisions made through the choice operator, G, use the emotion-like utility function J or J'. These decisions require information (J, U, Θ, f) that is produced outside of the choice agent itself, but which is inherited from a parent choice agent. Put in this context, values used in decision making become local instantiations of higher-order ob-

---

[2] This relationship between the past and the present is established through some learning process, be it evolutionary, reinforcement, or supervised.

[3] There is some debate about whether emotion precedes action or vice versa. This "chicken and egg" problem is resolved in the Peircean context wherein emotion can be both second of one process and firstness of another.

jectives. Associating objective parental values (secondness) with subjective child emotions (firstness) we reach the conclusion that *one role of emotions in intelligent choice is to place locally predicted or locally anticipated consequences in a global perspective using evaluations of past experiences.* Using the mathematical symbols developed herein, this is equivalent to saying that often the consequence operator f and the preference operator J, which exist in the parent (i.e., the evaluating or anticipating agent), are folded together in an emotion J' which exists in the choice agent. Thus, choice operates[4] on (J', U, Θ).

There are some interesting corollaries of this postulate. First, it means that *emotions are essential for decision making*. Since, unless the entire future can be known entirely, value judgments are inherently emotional, and since "[without] value judgment to support decision making, nothing can be intelligent, be it biological or artificial" [1], it is apparent that emotions are essential for intelligent choice. In terms of the discussion presented herein, this is because principles of rationality assume the role of thirdness wherein the firstness of emotions and observations are transformed into the secondness of decisions; rationality cannot interpret consequences nor generate decision without emotion. Second, it means that emotions are an efficient way of incorporating global objectives into a local decision. By allowing information about expected global consequences to be encoded in locally instantiated emotions, we effectually allow the global context of the problem (including past experience) to influence the decision. This occurs without using perfect causative models to predict future consequences.

A related hypothesis is that, contrary to the common perception, emotional responses are less subject to whim than action selection precisely because emotions change with slower temporal dynamics than rational choice (i.e., they are the products of higher order agents). This results from the differing temporal dynamics existent in multi-resolutional choice agents whence emotions, which are Peircean secondness for a higher-order agent, change more slowly than choices, which are Peircean secondness for a lower-order agent (it takes longer to modify emotions than it does to modify emotion-free thinking). If such a corollary holds true, then we may find that, under certain circumstances, rationality is a capricious cousin of emotion.

4 Opposing Emotions and Admissible Choice

We now turn attention to how opposing meta-values can be used to determine a subset of U that are admissible for rational choice. Recall that the conclusion from the previous section is that emotions are produced by higher order choices. These emotions manifest themselves in two ways. First and most obviously, they explicitly manifest themselves in the evaluation of consequences through J'. Second, they manifest themselves in a subset of admissible options (related to U). The first manifestation is well motivated in the previous section by associating emotions with value judgments so we will not address this point further. The second manifestation deserves further attention, and is addressed in this section.

Recall that anticipation and evaluation assume the role of meta choice agents with respect to the moment of truth. Unless anticipation and evaluation are simply re-enactments of the moment of truth, such meta decisions should not determine the precise action that should be taken under every possible circumstance, but rather what characteristics the decision should possess. Thus, meta decisions yield decision rules, and these decision rules permit sets of justifiable possibilities. With sets of justifiable possibilities as a background, some philosophers distinguish between categories of possibility in evaluating options. We adapt the categories put forth by Levi [7, 8]: *logical possibilities*, *serious possibilities*, *feasible possibilities*, and *admissible choices*. If U were restricted to the set of *logical possibilities* then the agent would only consider an option that is logically consistent with the agent's state of knowledge. *Serious possibilities* preclude many logical possibilities by eliminating options for which an alternative option is superior. *Feasible possibilities* preclude many logical possibilities by requiring an option to produce good enough consequences. Finally, *admissible possibilities* are both serious and feasible.

Let us make these definitions more precise. Let U denote a set of propositions closed under negation (that is, if u is in the set then so is the negation, $\bar{u}$). In determining a set of options for consideration, there must exist some criterion for eliminating a portion from the set of all logical possibilities. Given that the values used in making a decision are inherited from a parental choice problem, it is reasonable to assume that there are, from a local perspective, some reasons why an option should be selected and other reasons why an option should not be selected. In fact, there exists evidence suggesting that emotions are motivated by two opponent systems, attractive and aversive [6]. We need to translate these "pros" and "cons" into a decision rule that restricts the set of all logical possibilities to the set of feasible possibilities and to the set of serious possibilities. We formulate this process as a meta-decision problem ("meta" in the sense that we al-

---

[4] An unresolved problem, which is an area of future research, is to determine the tradeoff between information about future consequences encoded in the consequence operator f and encoded in the value function J'. In this paper, we restrict attention to the problem where meta-agents can use f but where choice agents use only J'-type information.

low a parent agent to perform the analysis to produce a rule that can be used by the child agent). The parent agent produces two independent value functions in the spirit of J': an attractive payoff for selecting option u given θ, $J_1(u,\theta)$, and an aversive payoff for rejecting option u given θ, $J_2(\bar{u},\theta)$.

4.1 Feasible Possibilities

We first consider the selection of feasible possibilities. Let $\phi : U \to \mathcal{B}$, where $\mathcal{B}$ is a sigma-algebra associated with U, denote a decision rule that maps the set of all logical possibilities U to the set of feasible possibilities given the state of nature. We can identify the utility of the decision rule ϕ as an aggregation of the two payoffs J_1 and J_2

$$\begin{aligned} J(\phi,x) &= E_{\theta|x}[J(\phi(\theta),\theta)] \\ &= E_{\theta|x}\left\{ \sum_{u\in U} \Big[\alpha J_1(u,\theta)p(\phi(\theta)=u|\theta) \right. \\ &\qquad \left. +(1-\alpha)J_2(\bar{u},\theta)p(\phi(\theta)=\bar{u}|\theta)\Big] \right\} \qquad (2) \end{aligned}$$

where α denotes a tradeoff parameter (resolved at the local perspective), and where $E_{\theta|x}(J)$ denotes the expectation of J with respect to the conditional probability $p(\theta|x)$, that is $E_{\theta|x}(J(\phi(\theta),\theta) = \sum_{\theta\in\Theta} J(\phi(\theta),\theta)p(\theta|x)$. Our objective is to find[5] a decision rule ϕ that maximizes this payoff keeping in mind that the rule ϕ determines which options are feasible possibilities given the current state. For convenience, define

$$\begin{aligned} \mu_A(u;x) &= \sum_{\theta\in\Theta} J_1(u,\theta)p(\theta|x), \\ \mu_L(u;x) &= \sum_{\theta\in\Theta} J_2(\bar{u},\theta)p(\theta|x). \end{aligned}$$

Maximizing (2) yields the decision rule

$$\phi(x) = \begin{cases} u & \alpha\mu_A(u;x) \geq (1-\alpha)\mu_L(u;x) \\ \bar{u} & \text{otherwise} \end{cases} . \qquad (3)$$

The set of feasible possibilities is thus given by

$$S_\alpha(x) = \{u : \alpha\mu_A(u;x) \geq (1-\alpha)\mu_L(u;x)\}. \qquad (4)$$

From (4) we note that the positive (attractive) motivation μ_A is compared to the negative (aversive) motivation μ_L to determine the set of feasible possibilities; the opposition between positive and negative decision forces determines feasibility. The local agent can therefore use the comparison rule in (4) to determine feasible possibilities. In words, only when an agent feels good enough[6] about an option can we consider it as a feasible possibility. *Feasible possibility is determined by comparing the benefit of an option (as manifest in a positive emotion) against the cost of the option (as manifest in a negative emotion) and permitting only those which are more beneficial than costly.*

4.2 Serious Possibilities

Shifting to serious possibilities, we observe that, according to the principle of maximum expected utility, a local choice agent will seek to select an option that maximizes a payoff. The payoff for an individual option u is equal to the payoff for accepting it minus the payoff for rejecting it whence the emotional payoff function can be defined as

$$J_\alpha(u;x) = \alpha\mu_A(u;x) - (1-\alpha)\mu_L(u;x).$$

The standard for serious possibility is obtained by defining the set of maximizing options as α varies

$$\mathcal{E}(x) = \{u : \exists\alpha \in [0,1] \text{ such that } u = \arg\max_{v\in U} J_\alpha(v,x)\}. \qquad (5)$$

Serious possibility is determined by comparing an option against all other logically possible options and retaining only those options that can possibly maximize emotional payoff. Because maximization over uncertain α is equivalent to domination[7] [5], serious possibilities can be locally determined by the rule that all dominated options should be eliminated.

4.3 Admissible Possibilities

The set of admissible options is determined as the set of logical options that are both feasible possibilities and serious possibilities $\mathcal{S}_\alpha = S_\alpha \cap \mathcal{E}$. The child agent can the use any criteria it likes in its choice given the parameters $u^* = G(J_\alpha, \mathcal{S}_\alpha, \Theta)$. A rational decision-maker uses positive and negative emotions to determine the set of possibilities that it can choose. *The set of admissible possibilities is determined by the meta-analysis of a parent agent to produce (Peircean secondness) a rule that can be used by the child agent to transform (Peircean thirdness) observations into options.*

4.4 Discussion

Deciding if an option is a feasible possibility requires very little information; only the ratio of attractiveness and aversion need to be considered. If an option is more attractive than aversive, that is $\mu_A(u;\theta) > \mu_L(u;\theta)$, then the option is feasible. This test does not require a complete preference pattern defined over all possible options, nor does it require any other option to even be identified. Options are evaluated on their own merits (in terms of emotional evaluations of perceived consequences relative to the context established by the global goal) without comparison to other

---

[5] By choosing the decision rule ϕ we also choose the probabilities $p(\bar{u} \in \phi(\theta))$ and $p(u \in \phi(\theta))$.

[6] This recalls the approach to a decision where an agent tallies all the pros of a decision and all the cons of a decision; if the pros outnumber the cons then a decision is a serious contender.

[7] Loosely speaking, an option is dominated if another option exists which has both higher μ_A and lower μ_L.

options. Thus, feasibility requires only limited information but, since it is a consequence-based rule reflecting the generation of good enough consequences, it is a mandatory requirement in intelligent choice. Feasibility is a function of emotional valence (meaning degree of attractiveness an option possesses in efficiently achieving a behavioral goal).

By contrast, an option is a serious possibility if, given two completely specified preference patterns (or a single, completely specified, aggregated preference pattern reflecting current emotional arousal), the option is not inferior to any other option. This requires not only an unambiguous preference pattern but also the identification of every other possible option. Thus, seriousness requires a great deal of information and, in situations of bounded rationality, is optional. Seriousness is a function of emotional arousal, and operates on the set of feasible possibilities to "break ties" between multiple options in S_α.

5 Conclusions and Future Work

Emotion is a function of not only the evaluations of past choices (and their resultant consequences), but also the expectations of future choices (and their anticipated consequences). Through past experience, emotion serves to place present circumstances into a global perspective and thus permits selection of a decision that produces desirable future consequences. Thus, emotion is a local instantiation of global objectives and, in the absence or certain knowledge of all future consequences, is a necessary companion of rational choice.

By exploiting the dichotomous motivational factors that can produce emotions, the set of all logical possibilities can be restricted to the set of admissible possibilities. This permits the decision-maker to make a choice from the set of feasible and, when possible, serious possibilities. Thus, emotion frames the choice problem and restricts attention to a managable set of options.

Future work in this area includes designing robots with opposing emotions, and using these opposing emotions to select feasible behaviors. Such a design should include a notion of emotional state which changes with success and failure under either a supervised or reinforcement learning paradigm. Furthermore, the robot must reach an appropriate resolution between information encoded in emotions (such as Q-functions from reinforcement learning) and information encoded in causative models.

A second area of future work is in modeling automobile driver behavior. There is evidence that opposing motivational factors can be used to determine how drivers will respond to certain circumstances [3, 4]. Experiments in this area will continue, and will be augmented by experiments designed to detect the modulatory influence of emotional arousal.

References

[1] J. S. Albus. Outline for a theory of intelligence. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3):473–509, May/June 1991.

[2] Antonio R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. G. P. Putnam, 1994.

[3] M. A. Goodrich and E. R. Boer. Semiotics and mental models: Modeling automobile driver behavior. In *Joint Conference on Science and Technology of Intelligent Systems ISIC/CIRA/ISAS'98 Proceedings*, Gaithersburg, MD, September 1998.

[4] M. A. Goodrich, E. R. Boer, and H. Inoue. A model of human brake initiation behavior with implications for ACC design. In *ITSC 99*, 1999. To appear.

[5] M. A. Goodrich, W. C. Stirling, and R. L. Frost. Model predictive satisficing fuzzy logic control. Technical Report CBR TR 98-3, Nissan Cambridge Basic Research, Nissan Research and Development, Inc., Cambridge, MA USA, 1998.

[6] P. J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, May 1995.

[7] I. Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, Massachusetts, 1980.

[8] Isaac Levi. *The Covenant of Reason: Rationality and the Commitments of Thought*. Cambridge University Press, 1997.

[9] M. Minsky. *The Society of Mind*. Simon and Schuster, 1986.