

# Counterfactual thinking and emotions: regret and envy learning

Giorgio Coricelli<sup>1</sup> and Aldo Rustichini<sup>2,\*</sup>

<sup>1</sup>*Institut des Sciences Cognitives, Centre de Neurosciences Cognitives, CNRS UMR5229, Université Lyon1, 67, Blv. Pinel 69675 Bron, France*

<sup>2</sup>*Department of Economics, University of Minnesota, 1925 4th Street South, 4-101 Hanson Hall, Minneapolis, MN, 55455-0462, USA*

Emotions like regret and envy share a common origin: they are motivated by the counterfactual thinking of what would have happened had we made a different choice. When we contemplate the outcome of a choice we made, we may use the information on the outcome of a choice we did not make. Regret is the purely private comparison between two choices that we could have taken, envy adds to this the information on outcome of choices of others. However, envy has a distinct social component, in that it adds the change in the social ranking that follows a difference in the outcomes. We study the theoretical foundation and the experimental test of this view.

**Keywords:** neuroeconomics; emotions; learning

## 1. REGRET AND ENVY IN CHOICE UNDER UNCERTAINTY

Ask a group of subjects to choose between two options: one is to take an amount of £20 to be paid for sure. The other option is a payment of either £100 or nothing, to be decided on the toss of a coin. We will call this second option a lottery.

### (a) *Regret*

A majority of the subjects will probably choose the random payment. Let us focus our attention on these subjects. When the coin is tossed, approximately half of them will get nothing, and will experience regret for the choice they made. The other half will get £100, and they all will congratulate themselves with their choice. These negative and positive affective states are puzzling: there is nothing different, from an ex ante point of view, in the choice made by subjects who won and those who did not. They had the same set of options, and they had made the same choice. Still, these emotions are experienced (Zeelenberg *et al.* 1996; Zeelenberg & van Dijk 2005), and the effect is not limited to the laboratory environment (Zeelenberg & Pieters 2004).

Also, since the probability of the outcome was fully specified, no new information is provided to subjects with the toss of the coin. In spite of this, if subjects were asked to make similar choices again, the outcome of the previous choices would probably affect the following decisions (Coricelli *et al.* 2005).

### (b) *Envy*

Suppose now that you take a similar sample of subjects and you randomly match them in pairs. Then you ask each subject to choose between lottery and certain

amount, and you inform them that the other subject is also choosing between the two same options. Let us focus on pairs of subjects who chose a different option, one the lottery and the other the certain amount. The subject who picked the lottery, if the other won, will be envious of the choice of the other. Similarly, if the subject who chose the lottery won, the subject who chose the certain amount will be envious. This response too is not limited to the laboratory experiment (Zeelenberg *et al.* 1996; Luttmer 2005).

In this case as well, the affective response of the subjects is puzzling: the only difference between the situation in the ex ante choice and the ex post evaluation is due to a random outcome that was clearly anticipated at the moment of choice, and in known and precise proportions. Presumably the subjects had taken this information into account at the moment of choice.

### (c) *Outline of the paper*

This paper will report on research linking these two emotional responses, the private ones (like regret) and the social ones (like envy).

It will proceed in two parts. In the first we will review experimental evidence that tests how these affective responses originate in a controlled laboratory environment. This will allow us to test and measure the effects that we want to study, and motivate the analysis developed in the second part.

As we noted, these affective responses and the impact of later choices are puzzling from the point of view of a rational evaluation of one's choices. So in the second part of this paper we will examine an explanation of this puzzle, by analysing the functional role that these emotions have in learning. An analysis that is testable, however, requires a precise model and precise quantitative predictions. Our contribution here is to outline the conditions under

\* Author for correspondence (aldo.rustichini@gmail.com).

One contribution of 12 to a Theme Issue 'Rationality and emotions'.

which this role is effective, and some problems that are still open. To do this, we need to develop a model of adaptive learning, and then consider the consequence on the learning process of introducing the counterfactual thinking, that is thinking about what might have been (Lewis 1973; Olson 1995; Byrne 2002).

#### (d) *Emotions, rationality and learning*

Let us begin with the main idea that is going to be developed here. Envy and regret share a common feature, the counterfactual analysis of the individual's actions. In the experiments we have just described, a subject evaluates the outcome of a choice he made by comparing it with the outcome of choices he did not make. That is, he considers what might have been (so he does a counterfactual analysis) had he chosen a different action (so he focuses his analysis on his own personal responsibility). Regret considers actions we could have taken, but did not take, and for which we get to know the outcome. Envy considers actions that we could have taken, we did not take but someone else did, and for which we get to know the outcome that the other person obtained.

Both counterfactual analysis and personal responsibility are essential. If the outcome of the lottery is the low payment, a subject may also compare what he received with what he would have received had the outcome been different, that is, had Nature chosen a different outcome. In this case he will experience disappointment: he is still using counterfactual thinking, but applied to the role of Nature, not his own. Similarly, an individual may experience a negative affect because of the outcome of others that was beyond his reach: for example he may be envious of the height of someone else.

This view is very close to the one put forward by Festinger in his theory of social comparison processes (Festinger 1954; Suls *et al.* 2002). He proceeds from the very reasonable assumption that individuals have a drive to evaluate accurately their own abilities. How can they give an accurate evaluation? Typically, they may try to use, when they are available, objective measurements of their performance. What can they do if these means are not available or are unreliable? In his second hypothesis, Festinger (1954, p.118) postulates that when objective, non-social means are not available, then people evaluate their abilities by comparison with the abilities of others.

We take here a similar point of view: both regret and envy have functional role, the one of helping the individual in learning to give an evaluation of the actions he has available in light of his past experience. We develop this idea in §4, and we examine some of the problems that are open in §5.

## 2. EXPERIMENTAL TESTS

A controlled experimental test of the affective response of regret and relief in choice under uncertainty is provided in Mellers *et al.* (1999) and Coricelli *et al.* (2005).

#### (a) *Regret and disappointment*

The experiment is designed to test the differential effect of counterfactual evaluation of the consequences of different random outcomes for a given choice (the effect of nature's choices) as opposed to the evaluation of the consequences of the individual's action.

Subjects had to make choices in several trials. In each trial, the subject had to choose between two lotteries displayed on a computer screen. The probability of each outcome was described as a sector on a circle, and the subjects were informed that every point on the circle had equal probability.

After the subject had made his choice, a square framed the lottery he had chosen, to remind him of the choice he had made. The display of the other lottery was kept on the screen. Then a spinner spun on both circles, and stopped randomly at some point, indicating the outcome. Since this happened on both lotteries, the subject knew the outcome of both. He was then asked to rate how he felt about the outcome, on a fixed scale symmetric around the zero. Regret was defined the event in which the outcome for the chosen lottery is smaller than the outcome on the other lottery, and relief the event in which the opposite happened.

A control condition is provided by trials where the non-chosen lottery is hidden after the subject's decision, and only the outcome of the chosen lottery was kept on the screen. In these trials the only comparison subjects could make was the one between the realized outcome and the alternative, non-realized outcome of the chosen lottery. In this case disappointment (and, respectively, elation) was defined as the event in which the realized outcome was smaller (larger) than the alternative outcome.

In Camille *et al.* (2004) this design was used to test the difference in response between normal subjects and patients with lesions in the orbito-frontal cortex (OFC) in conditions of disappointment and regret. OFC patients were found to have behavioural responses similar to normal subjects in disappointment, but completely different in regret: one can take the difference between the amount won in the chosen lottery and the amount won in the unchosen lottery as a measure of the potential for subjective feeling of regret. The subjective evaluation given by normal subjects was very sensitive to this difference but it was not in the case of OFC patients. That is, OFC patients, while able to code disappointment, seemed to be unable to code emotion like regret that codes the personal responsibility of an outcome.

The same design was used in Coricelli *et al.* (2005) with normal subjects to detect patterns of brains activation in conditions of regret and relief, and contrast them with trials in which disappointment was experienced. The OFC was found to code the emotional response of regret. In Coricelli *et al.* (2005) the authors reported that, across their fMRI experiment subjects became increasingly regret aversive, a cumulative effect reflected in enhanced activity within OFC and amygdala. Under these circumstances the same pattern of activity that was expressed with the experience of regret was also expressed just prior to choice, suggesting the same neural circuitry mediates

both direct experience of regret and its anticipation. Thus, the OFC and the amygdala contribute to learning based on past emotional experience.

### 3. REGRET AND ENVY

The experiments we have just described give a simple and effective tool to test the hypothesis that envy is just the social correspondent of regret. This experiment is reported in Bault *et al.* (2008). Subjects participated in the experiment in pairs that were randomly created and called to the laboratory. The experimental design emphasized the similarity between envy and regret, using two conditions: a one-player condition to test the effect of regret and a two-players condition to test the effect of envy. The one-player condition was identical to the experimental design described earlier. The two-players condition was very similar to the one-player, but after his choice, the subject observes the choice that a subject like him has made out of the same two options available. If the two subjects had chosen the same lottery and had the same outcome, then they will experience what we can call shared regret or shared relief. If they choose a different lottery, then they might experience envy (if their outcome is lower than the outcome of the other) or gloating (if the opposite occurs). In the experiment, subjects were facing choices made by a computer program.

Consider now our initial hypothesis that learning is just social regret. If this hypothesis is correct, then there should be no substantial difference in ratings in the two conditions for any given pair of outcomes of the chosen and unchosen lottery. In Bault *et al.* (2008) the authors measured, in addition to self-reported emotional evaluations, the skin conductance response (SCR) of the subjects: this is a measure of electrical conductance of the skin, and indirectly of the level of emotional arousal of the subject. For this measurement one should not expect any difference at the moment in which the outcome of the two lotteries is displayed. Non-parametric tests are used to check the significance of the difference.

For negative emotions, envy was stronger than regret: the average scores in the affective scale ranging from  $-50$  (extremely negative) to  $+50$  (extremely positive) were  $-29.19$  and  $-25.27$ , respectively, with a value  $z = 2.754$ , and  $p = 0.0059$ . Also regret was stronger than shared regret (shared regret had an average score of  $-18.49$ ,  $z = 4.120$ ,  $p = 0.00001$ ). For positive emotions, gloating was stronger than relief (with score  $33.04$  and  $25.62$  respectively,  $z = 4.032$ ,  $p = 0.0001$ ) and relief was stronger than shared relief (shared relief has score  $19.91$ ,  $z = 4.620$ ,  $p = 0.00001$ ). SCR correlated with the self-emotional ratings ( $r = 0.93$ ,  $p = 0.006$ ); moreover, the magnitude of SCR in the two players condition for different choices was higher than in the one player condition. In summary, the two player emotions when the subjects made a different choice are stronger than the single player ones. In particular, gloating, or the joy of winning, was stronger than relief. Clearly subjects liked inequality, as long as they were at the top of the scale. This finding seems to contradict the hypothesis

that individuals are, in general, better off when the distance between them and individuals with inferior outcomes is reduced, at least when their outcome is not changed, and perhaps even when it is (Ernst & Schmidt 1999; Ernst & Fischbacher 2002).

#### (a) Learning and social evaluations

Two conclusions seem clear. The first is that envy and regret, as well as their positive counterparts, share the common nature that is hypothesized in the functional role explanation: they are affective responses to the counterfactual evaluation of what we could have gotten had we made a different choice. Envy has, like regret, a functional explanation in adaptive learning.

The results also show that the social emotions have an additional role, since the response that they evoke is more powerful. In other words, envy is likely to be the resultant of two distinct components: one is driven by learning the consequences of one's actions, and is closely related to regret. The other is a measure of one's ranking in a social scale, and is profoundly different from regret. In fact, it can arise even when the reason for the dissatisfaction is not our own responsibility (as when, for example, we envy someone's height).

Of course this describes the typical, or average, response of individuals. An interesting aspect of the analysis is given by the individual differences with respect to regret and relief. An axiomatic analysis provides the basis for this extension to individual characteristics (Maccheroni *et al.* 2008).

We can now proceed to examine more closely the functional explanation, in a precise model.

### 4. ADAPTIVE LEARNING AND COUNTERFACTUAL EVALUATION

We plan to develop here a model of adaptive learning where the observation of the outcome of the unchosen options improves the decisions taken in the learning process.

The problem we consider is classical: an individual has to make choices over infinitely many periods. Before he decides, he observes a current state, chooses an action out of a feasible set, collects a reward for that period and goes to the next period. In the new period a new state is determined, and the entire procedure is repeated. Future rewards are discounted.

The set of states is called  $S$ , with a generic element  $s$ . An individual chooses an action  $a$  out of a set  $A$  of feasible actions. For simplicity, and without loss of generality, this set is the same in every period and is independent of the state. Both sets are finite.

For a given pair of states and action the individual receives a reward  $r$ . The rewards are not deterministic. To illustrate, consider the introductory example as an example of the problem we are analysing, a choice of the sure amount delivers a payment of certain quantity, but the choice of the lottery only gives us a probability over outcomes. This randomness of the outcomes associated with our choice is an important and realistic feature of our real life choices: many events which are outside our control influence the

outcome of our decisions, from the education we choose, to the investment we make, down to the choice of the means of transportation for the day.

Actions do not only influence rewards, but also affect what the future state will be. For any given pair of state and action, there is a probability to transit to a new state in the next period. This is a key feature of the problem: a good choice must not only take into account the current reward, but also the effect on the transition to the future state. This is also a common feature of real life problems, where what we do today affects not only our rewards today, but also what will happen tomorrow. For example, the choice of one college degree over another has a strong influence on the states we will face in the future.

#### (a) *Information*

A crucial manipulation used in experimental tests described in §3 was the different information provided to the individual. To illustrate these different conditions, consider a problem where states are sets of lotteries. This is the situation in our introductory example, where a state is a pair of a lottery and a certain amount.

In the *incomplete feedback* condition the individual who in state  $s$  has chosen some action  $a$  is informed only of the outcome of the random variable  $r(s,a)$ , that is of the lottery he has chosen. In the *complete feedback* condition he is informed of the outcome of all the lotteries, those he chose and those he did not. This manipulation allows the experimenter to separate the behavioural and brain correlates of the comparison between disappointment and regret.

#### (b) *The value function*

A benchmark for this problem is the value to the decision maker when he uses an optimal policy. The optimal policy defines for every initial state the sequence of choices that the individual has to make in every period, taking into account the past history of actions and states, if he wants to maximize future discounted rewards. The value function describes for every initial state the infinite discounted expected reward in the future under this policy. This value function is unique, and the optimal choice in every period is only dependent on the state, and does not need to look at the full previous history of actions and states.

#### (c) *Adaptive learning*

Theories of adaptive learning are efforts to explain how the optimal solution to the problem of sequential choice can be learned as the outcome of a process that adjusts the current value function.

In an adaptive learning formulation of the problem the learner does not know the two functions that describe the reward and the transition, and does not even attempt to learn them. A good adaptive model has to satisfy the requirement that the sequence of choices converges to the optimal solution, no matter what the reward function and the transition functions are.

In adaptive learning, the function  $V$  is approximated by a sequence  $V_k$  in every stage of the approximation. In every period, for the given pair of

states and actions, the new value function is given by an incremental adjustment of the function obtained in a previous stage. The adjustment is proportional to the prediction error, which is the difference between the expected and the realized value for that period. Both expected and realized values are far-sighted, in that not only the current reward but also the continuation value from the next state are taken into account.

A similar adjustment is possible for the choice of action. In every period, each action is chosen with some probability that can be changed depending on past rewards obtained. Once the current reward is obtained from an action, the probability in the next period of choosing that action can be increased by a factor proportional to the reward obtained with that action.

#### (d) *Full and partial information*

Typically, the adjustment to a new value function only uses the information provided on the outcome of the chosen action (Schultz *et al.* 1997; Schultz 2002). For such an individual, the adaptive process described has an obvious shortcoming: the information that he has on the outcomes of actions different from the one he chose is not used in the learning process. This is of course not the case for a decision maker who computes the optimal solution because this decision maker *knows* the function  $r$ , so he uses the knowledge of the function to compute the value and the optimal policy. For an adaptive learner, ignoring the outcomes of actions different from the one he chose means ignoring important information on the function  $r$  that is provided by the outcomes of the actions in the set  $A$ . It is clear, intuitively, that the use of this information should be part of the real learning process that we observe. The problem is: how is this knowledge incorporated into an adaptive learning process?

The answer to this question may be very simple or very complex, depending on the environment we consider. The fundamental distinction hinges on a property of the transition function  $T$ . If the action taken in the current period does not influence the realization of the state in the next period, then the problem we are studying is considerably simpler. Instead, in the more general and more interesting case in which actions do affect the transition to the new state, an important problem arises (see §5). Let us begin from the simpler case, for which the answer is known.

#### (e) *Regret learning*

The theory we adopt makes reference to existing theories of regret as a form of adaptive learning, in the tradition of the Megiddo–Foster–Vohra–Hart–MasColell (Megiddo 1980; Foster & Vohra 1999; Hart & Mas-Colell 2000; Foster & Young 2003; Hart 2005) regret-based models. In these theories, learning adjusts the probability of choosing an action depending on the difference between the total rewards that could have been obtained with the choice of that action and the realized total rewards.

For example, in the Hart–MasColell model the regret for having chosen the action  $a$  instead of  $b$  is the positive part of the difference between the total

reward obtained if action  $b$  had been chosen instead of  $a$  in the past, and the total value that has been realized with the actions really chosen. That is, we compute the difference between the two values: if this difference is positive, then that is the regret; if the difference is negative, no regret is assigned. The probability of choosing an action in the next period is then determined in two steps as follows. First, we determine whether the action should be changed. If action  $a$  was chosen in the previous period, then the probability of choosing a different action in the current period is proportional to the total regret over the actions different from  $a$ . If the decision is to pick a different action, then we need a second step to decide which action we switch to. The probability that an action  $b$  is chosen is again proportional to the amount of regret for having chosen in the past  $a$  instead of  $b$ . Consider this procedure in terms of the general model we have described in the previous section: in that model a different state presents a different set of options. Since in the current case the set of options is the same in every period, it is clear that we are considering the case in which there is a single state, that is  $S \equiv \{s\}$ . This procedure has good optimality properties: the Megiddo theorem for the single player case, and the Foster–Vohra–Hart–MasColell theorems for games show that this procedure converges to optimal choices in the single player case and to correlated equilibria in the case of games.

In the literature on machine learning the difference between the two conditions (full and partial feedback on the actions) has been explored, and some results highlight the advantage induced in the full feedback condition. For example, in Auer *et al.* (1995, 2002) the authors examine the loss to a decision maker who has to choose one action out of a set; his payoff depends on his choice of action and the choice of an adversarial opponent, who is not constrained in any way in the choice of the action. The measure of performance they use is the difference between the maximum the decision maker could have achieved *ex post*, given the payoff that the opponent has assigned to the different actions, and the average payoff actually realized. They show that the loss per unit of time from the maximum that can be achieved is of the order  $O(T^{-1/2})$  of in the length  $T$  of the problem in the full feedback condition and  $O(T^{-1/3})$  in the partial feedback condition. Numerical simulations of behaviour of neural networks in Marchiori & Warglien (2008) show that the introduction of regret in the feedback improves substantially the performance of the network. Of course since the results are obtained by numerical simulation the results are harder to interpret; analytical results would be important.

#### (f) *Regret learning and prediction error*

The idea that learning may use regret, that is the comparison between what the chosen action gave and what other actions might have given can be introduced into the learning based on prediction error, both at the stage in which the new value function is evaluated and at the stage in which the probability on actions is chosen.

For example, we have seen that the probability of choosing a certain action in the next period can be updated by considering only the outcome of the chosen action, by increasing the probability of choosing it next time proportionally to the reward obtained with that action. But when the reward of all the actions is available, a more effective adjustment is possible. For example, the probability can be increased by an amount proportional to the difference between the reward from that action and the maximum that is obtained from the other actions. This difference is measured by the regret experienced by the individual, and makes the adjustment more effective. A similar modification can be made of the process adjusting the value function.

### 5. THE ATTRIBUTION PROBLEM

What we have concluded so far, however, depends crucially on our initial assumption that the action affects rewards but not the transition to the state in the next period. This is clearly an important case but very limited, also ecologically. We can now return to the analysis of the complex and more interesting case in which the state depends on the action of the individual.

A first difficulty in establishing an analytic foundation for the role of regret is that the effect must be established quantitatively and qualitatively. Consider, for example, models of  $Q$ -learning (Watkins 1989). In these models a vector representing the current approximation to the true value is updated in every period by an amount proportional to the prediction error, ignoring the information on the payoff of the other actions. Under some mild technical conditions the process converges to the  $Q$ -value that is obtained by following the optimal policy after the first period (Watkins & Dayan 1992). It follows that the improvement that can be introduced by considering, through regret, the payoff of the other actions, cannot consist in a better limit behaviour, since the optimal one can already be obtained ignoring the payoff of the non-chosen actions. An improvement in a different dimension can be introduced: for example, one may show that regret induces a faster convergence to the optimal solution, or a smaller loss in the trajectory leading to the limit. So even if the limit is the same, the speed of convergence to it is faster.

A more fundamental problem in any theory of adaptive learning that introduces counterfactual thinking into the analysis of the learning process is what we can call the attribution problem. The problem is easy to understand.

The choice of the current action determines the current reward, and also the transition to the next state. Both effects influence the value at the state. Consider the action prescribed by the optimal choice. It may be the case that the reward for that period which can be obtained from a different action is higher. In spite of this, of course, the action prescribed by the policy may still be optimal, because the action with the higher current payoff may induce a transition to a ‘bad’ state with low payoffs.

To illustrate the problem, consider the simple case in which there are two states, one 'good' and one 'bad'  $S = \{G, B\}$  and two actions, say  $P$  (=Prudent) and  $M$  (=Myopic). The rewards in the first state are 5 for the  $M$  action, and only 1 for the  $P$  action; the rewards in the second state are 0 for both actions. The probability that the state changes for sure from the state  $G$  to the state  $B$  is 1 with the  $M$  action, while with the action  $P$  the state stays  $G$  for sure. Instead, for both actions the probability of reverting from the second (with zero rewards) to the first is very low. It is clear that the action  $P$  is optimal, in view of the fact that it maintains the 'good' state. But the comparison, made at the optimal policy, between  $P$  and  $M$  is in every period unfavourable to the action  $P$ , which gives a lower payoff all the times. The difficulty, of course, is that the transition to the bad state  $B$  that the choice of action  $M$  would induce is not observed. The regret would induce in this example the wrong action.

#### (a) *Social and private learning*

A good solution of the problem of integrating regret into adaptive learning when the transition among states depends on the action of the individual is still not available. The fundamental difficulty is in the asymmetry of the information available on the consequences of actions taken: even in the full feedback condition, in which the learner knows the rewards associated with the different actions, he does not know the effect of all the actions on the next state, because this effect is only observable for the action that was really chosen.

This difficulty, however, is only in the case of private learning. When we observe others taking action, we can also observe the separate effect that *their* action has on *their* state. In social learning, the integration of counterfactual thinking into learning is easy, because both effects (on reward and on state transition) of all actions are observed.

## 6. CONCLUSIONS

The experimental results and the theoretical analysis we have reviewed suggest an adaptive role of emotions, like regret and envy, which have two distinguishing features. First they are based on rewards. Second, they proceed from a counterfactual consideration of outcomes.

Our analysis has emphasized how private and social emotions (like regret and envy) are closely related precisely because they both fulfil the role of effectively evaluating our past actions. It has also pointed out some important difference: the most important is probably that counterfactual evaluation is easier in social environments, because the effects on current rewards and those on future rewards (that in our model carry over through the state) can be separated.

Our analysis also puts the relationship between emotions and rational choice in a different light. A remarkable result in the theoretical literature that has studied regret as an adaptive emotion is that if players in a game minimize regret, then the frequency of their choices converges to a correlated equilibrium of the

game, which can be considered the rational behaviour of players in the strategic environment. This has a general implication for our understanding of the role of emotions in decision making. In particular, it rejects the view that emotion and cognition (or rationality) are in conflict, by showing the implications of full integration between those two components of human decision making. Within the formal and functional approach used here, emotions do not necessarily interfere with rational decision making, on the contrary, they may implement it: they are a way of evaluating past outcomes to adjust choices in the future.

These are features which are common between the prediction error model and the counterfactual learning. The crucial difference between models of temporal differences learning (e.g. Schultz *et al.* 1997) and regret learning is the counterfactual difference between the rewards the individual received and those he would have received had he chosen a different action. Both the relationship and the differences between the prediction error model and the counterfactual model are clear. One important difference is of course the neural basis of the two: from the existing literature on the topic we know that the ventral tegmental area and ventral striatum are usually associated with the prediction error, while counterfactual learning is associated with the OFC.

## REFERENCES

- Auer, P., Cesa-Bianchi, N., Freund, Y. & Schapire, R. 1995 Gambling in a rigged casino: the adversarial multi-armed bandit problem. *Proc. 36th Annu. Symp. on Foundations of Computer Science*, FOCS '95, 322–349. See <http://www.computer.org/portal/web/csdl/doi/10.1109/SFCS.1995.492488>.
- Auer, P., Cesa-Bianchi, N., Freund, Y. & Schapire, R. 2002 The non-stochastic multi-armed bandit problem. *SIAM J. Comput.* **32**, 48–77. (doi:10.1137/S0097539701398375)
- Bault, N., Coricelli, G. & Rustichini, A. 2008 Interdependent utilities: how social ranking affects choice behavior. *PLoS ONE* **3**, pe3477. (doi:10.1371/journal.pone.0003477)
- Byrne, R. M. 2002 Mental models and counterfactual thoughts about what might have been. *Trends Cogn. Sci.* **6**, 426–431. (doi:10.1016/S1364-6613(02)01974-5)
- Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J. R. *et al.* 2004 The involvement of the orbitofrontal cortex in the experience of regret. *Science* **304**, 1167–1170. (doi:10.1126/science.1094550)
- Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A. *et al.* 2005 Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* **8**, 1255–1262. (doi:10.1038/nn1514)
- Ernst, F. & Fischbacher, U. 2002 Why social preferences matter—the impact of non-selfish motives on competition, cooperation, and incentives. *Econ. J.* **112**, 1–33. (doi:10.1111/1468-0297.00027)
- Ernst, F. & Schmidt, K. M. 1999 A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868. (doi:10.1162/003355399556151)
- Festinger, L. 1954 A theory of social comparison processes. *Hum. Relat.* **7**, 117–140. (doi:10.1177/001872675400700202)
- Foster, D. P. & Vohra, R. 1999 Regret in the on-line decision problem. *Games Econ. Behav.* **29**, 7–35. (doi:10.1006/game.1999.0740)

- Foster, D. P. & Young, H. P. 2003 Learning, hypothesis testing, and nash equilibrium. *Games Econ. Behav.* **45**, 73–96. (doi:10.1016/S0899-8256(03)00025-3)
- Hart, S. 2005 Adaptive heuristics. *Econometrica* **73**, 1401–1430. (doi:10.1111/j.1468-0262.2005.00625.x)
- Hart, S. & Mas-Colell, A. 2000 A simple procedure leading to correlated equilibrium. *Econometrica* **68**, 1127–1150. (doi:10.1111/1468-0262.00153)
- Lewis, D. 1973 *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Luttmer, E. F. P. 2005 Neighbors as negatives: relative earnings and well-being. *Q. J. Econ.* **120**, 963–1002. (doi:10.1162/003355305774268255)
- Maccheroni, F., Marinacci, M. & Rustichini, A. 2008 Social decision theory: choosing within and between groups. Mimeo.
- Marchiori, D. & Warglien, M. 2008 Predicting human interactive learning by regret-driven neural networks. *Science* **319**, 1111–1113. (doi:10.1126/science.1151185)
- Megiddo, N. 1980 On repeated games with incomplete information played by non-bayesian players. *Int. J. Game Theory* **9**, 157–167. (doi:10.1007/BF01781370)
- Mellers, B. A. 2000 Choice and the relative pleasure of consequences. *Psychol. Bull.* **126**, 910–924. (doi:10.1037/0033-2909.126.6.910)
- Mellers, B. A., Schwartz, A. & Ritov, I. 1999 Emotion-based choice. *J. Exp. Psychol.* **128**, 1–14. (doi:10.1037/0096-3445.128.3.332)
- Olson, J. M. 1995 *What might have been: the social psychology of counterfactual thinking*. Mahwah, NJ: Erlbaum.
- Schultz, W. 2002 Getting formal with dopamine and reward. *Neuron* **36**, 241–263. (doi:10.1016/S0896-6273(02)00967-4)
- Schultz, W., Dayan, P. & Montague, P. R. 1997 A neural substrate of prediction and reward. *Science* **275**, 1593–1599. (doi:10.1126/science.275.5306.1593)
- Suls, J., Martin, R. & Wheeler, L. 2002 Social comparison: why, with whom and with what effect? *Curr. Dir. Psychol. Sci.* **11**, 159–163. (doi:10.1111/1467-8721.00191)
- Watkins, C. J. 1989 Learning from delayed rewards, PhD thesis, University of Cambridge, UK.
- Watkins, C. J. & Dayan, P. 1992 Q-learning. *Mach. Learn.* **8**, 279–292.
- Zeelenberg, M. & Pieters, R. 2004 Consequences of regret aversion in real life: the case of the Dutch postcode lottery. *Organ. Behav. Hum. Decis. Processes* **93**, 155–168. (doi:10.1016/j.obhdp.2003.10.001)
- Zeelenberg, M. & van Dijk, E. 2005 On the comparative nature of regret. In *The psychology of counterfactual thinking* (eds D. R. Mandel, D. J. Hilton & P. Catellani), pp. 147–161, Routledge Research International Series in Social Psychology (ed. P. Robinson). London, UK: Routledge.
- Zeelenberg, M., Beattie, J., van der Pligt, J. & de Vries, N. K. 1996 Consequences of regret aversion: effects of expected feedback on risky decision making. *Organ. Behav. Hum. Decis. Processes* **65**, 148–158. (doi:10.1006/obhd.1996.0013)