



ELSEVIER

European Economic Review 46 (2002) 633–655

EUROPEAN
ECONOMIC
REVIEW

www.elsevier.com/locate/econbase

Presidential Address

Rational irrationality: Some economics of self-management[☆]

Jean Tirole^{*,1}

*Institut D'Economie Industrielle, Manufacture des Tabacs, 21 Allées de Brienne,
31000 Toulouse, France*

Abstract

The paper first reviews some of the main challenges posed to our received economic paradigm by a variety of behaviors documented by psychologists and discusses possible responses to this challenge. It then describes a specific attempt at unifying a number of themes in social psychology using a parsimonious modeling of human behavior. © 2002 Published by Elsevier Science B.V.

JEL classification: D81; D84; D90

Keywords: Motivated beliefs; Self-confidence; Willpower; Rationality

1. Introduction

Exciting research is presently being performed in the area of psychology and economics. While economists have been interested in behavioral patterns for decades and even centuries, the current movement is attracting a large amount of (mainly young) economists and is characterized by sustained interaction with psychologists. The purpose of this lecture, and that of the Marshall and Schumpeter lectures delivered by Rabin (2002) and Fehr (2002) is to take stock of advances in some specific areas of psychology and economics and to provide perspectives on where the field might be going.

[☆] Presidential address, European Economic Association, delivered in Lausanne on August 30, 2001. I have been tremendously fortunate to collaborate with Roland Benabou on topics related to the content of this lecture. This is in many respects a joint lecture. The two of us have recently much benefitted from working with Marco Battaglini on this research agenda.

* Tel.: +33-1-61-12-86-42; fax: +33-5-61-12-86-37.

E-mail address: tirole@cict.fr (J. Tirole).

¹ IDEI and GREMAQ (CNRS UMR 5604), Toulouse, CERAS (CNRS URA 2036), and MIT.

Much economic analysis builds on a deliberately simple-minded description of human preferences and behavior, in which the individual is depicted as maximizing at each instant t over some action set A_t the expectation of the present discounted value of the flow utility of consumption $u_\tau(c_\tau)$ given the information I_t he has accumulated prior to date t :

$$\max_{A_t} E\left[\sum_{\tau \geq t} \delta^{\tau-t} u_\tau(c_\tau) \mid I_t\right].$$

Economists have of course long been aware of the crudeness of this representation; but they have argued that its parsimony is also its strength. Social scientists (be they economists, sociologists, psychologists, political scientists or others) should not content themselves with observing and portraying human behavior; they also have a social responsibility to suggest and guide policies that promote economic development, reduce inequality, make product and financial markets function better, improve macroeconomic performance, and so forth. To this end, they need a parsimonious framework that has predictive power and normative content.

Yet, psychologists have accumulated an impressive body of evidence suggesting ways in which we economists might want to reconsider and adjust our standard paradigm. Section 2 discusses the challenge created by these findings, and Section 3 the economists' possible responses to this challenge. Sections 4 through 6 describe an attempt at unifying a number of themes in social psychology using a parsimonious modeling of human behavior. Section 7 offers concluding remarks.

2. The challenge

Reviewing those findings in psychology that seem most relevant to economists is out of the scope of this lecture; there are several excellent surveys (e.g., Rabin, 1998). To make my main points, though, I need to remind the reader of some critiques of the homo economicus paradigm (my review will be highly selective and the citations somewhat random). This brief review will be organized around four categories of departures from the economic paradigm, calling into question its assumptions regarding, respectively, optimization (the act of maximizing over some well-defined action set A_t), judgment (the formation of beliefs and expectations given past information I_t), discounting (and its embodied time consistency assumption) and preferences (the narrow set of variables affecting flow utility).

I will proceed in reverse order. Thus, our first category of departures from the economic paradigm relates to the specification of experienced utility.

2.1. Preferences: $\max_{A_t} [\sum_{\tau \geq t} \delta^{\tau-t} u_\tau(c_\tau) \mid I_t]$

2.1.1. Adding nonstandard hedonic components

A number of authors have suggested that the vector of “goods” entering the flow utility should be enlarged so as to account for various psychic costs and benefits

experienced by the individual. For example:

- the anticipation of future consumption or savoring.² Illustrations of anticipatory feelings are the facts that one is in a good mood before a vacation or anxious before undergoing surgery. Savoring can be modeled for example by adding anticipated consumption $c_{\tau+1}^a, c_{\tau+2}^a, \dots$ (or higher moments thereof) into the flow payoff: $u_{\tau}(c_{\tau}, c_{\tau+1}^a, c_{\tau+2}^a, \dots)$,
- self-image, to the extent that the flow payoff $u_{\tau}(c_{\tau}, \hat{\theta})$ is conditioned by self-esteem $\hat{\theta}$, where the parameter $\hat{\theta}$ often stands for “ability” but may alternatively denote generosity, beauty or any other trait valued by the individual,³
- in a social context, the flow utility may depend on the consumption of other individuals, whether this internalization arises from a concern for fairness,⁴ from a quest for status, or from any other reason; for example, Fehr and Schmidt (1999) describe the flow utility of individual i in interaction with individual j as depending not only on individual i ’s consumption c_{τ}^i but also on the absolute difference of consumptions: $u_{\tau}^i(c_{\tau}^i, -|c_{\tau}^i - c_{\tau}^j|)$,
- and last the flow utility may account for emotions,⁵ such as anger, liking, reciprocal altruism, guilt, shame or pride. For example, I am more inclined to conserve water during a drought or pay my taxes when other people also do, even though this is precisely when my conserving water or my paying taxes has the lowest marginal social utility.

2.1.2. Reference points and path dependency

Another set of suggestions regarding flow utility is that this utility depends on a comparison with so-called reference points; for example the flow utility may take the form $u_{\tau}(c_{\tau} - \bar{c}_{\tau})$, where \bar{c}_{τ} denotes the date τ reference point.

To be certain, economists routinely account for habit formation and addiction, for example, by making marginal utilities depend on past consumptions;⁶ still we may insufficiently account for the fact that consumption gradients may matter as much as absolute levels. In models putting special weight on sensitivity to change, then $\bar{c}_{\tau} = c_{\tau-1}$. More generally, people often seem to assess utilities from consumption in comparison with a variety of reference levels. Particularly interesting puzzles in this respect include:

- loss aversion⁷ (the fact that people are more averse to losses than they are attracted to same-size gains, and this even for small risks),
- the related endowment effect,⁸ according to which mere ownership on average increases the willingness to pay for a good, or, more generally, according to which the

² E.g., Bentham (1970), Marshall (1881), Jevons (1905), Loewenstein (1987) and Caplin and Leahy (2001).

³ E.g., Akerlof and Dickens (1982), Köszegi (1999) and Weinberg (1999).

⁴ E.g., Fehr et al. (1993, 1998).

⁵ E.g., Akerlof (1984), Frank (1988) and Rabin (1993, 2002).

⁶ E.g., Becker and Murphy (1988) and Constantinides (1990).

⁷ E.g., Kahneman et al. (1990) and Rabin (2000b, 2002).

⁸ E.g., Thaler (1980).

willingness to pay for a good is smaller than the minimum compensation demanded for the loss of the same entitlement even when income effects are negligible (for example, in questionnaires, people demand much more for accepting a small increase in the probability of a sudden death than they are willing to pay for eliminating an existing and equal probability), and

- the sunk cost fallacy, the fact that if I lose my opera tickets, I may not want to purchase new ones even if the transaction costs have not gone up.⁹

2.1.3. *Stepping back for a moment*

At this stage, and before we have even discussed other ways in which the economic paradigm could be made more realistic, the reader probably is already bewildered. Clearly, we, as a profession, should not impulsively add a new element into the utility function every time we cannot readily explain a behavior or an apparent concern.

First, we would lose parsimony and thereby predictive power. As Bob Frank (2001) puts it:

The problem here is that if analysts are totally unconstrained in terms of the number of goals they can attribute to people, virtually any behavior can be “explained” after the fact simply by positing a taste for it. A man drinks the used crankcase oil from his car, then writhes in agony and dies moments later? No problem, if we are free to assume that he really *liked* crankcase oil. As students of the scientific method are quick to emphasize, a theory that can explain everything ends up explaining nothing at all. To be scientifically valuable, a theory must make predictions that are at least in principle capable of being falsified.

And hence the dilemma confronting proponents of rational choice theory: versions that assume narrow self-interest are clearly not descriptive, whereas those to which goals can be added without constraint lack real explanatory power.

Second, we would end up with potential contradictions. For example, the sensitivity to change suggests that the key to happiness is low expectations, an hypothesis that does not accord well with the benefits from savoring and/or with the demand for a positive self-image, which both suggest benefits from high expectations. These observations cannot be inconsistent, which they may end up being if one systematically adopts reduced form approaches.

Third, and relatedly, we may need to dig deeper into the real motivation of behaviors. Are the extra ingredients really part of the preferences or are they purely instrumental? For example, the demand for self-esteem may come from the experience of reflexive consciousness, as when you lie awake in bed at night or look in the mirror glorying in your triumphs or thinking about your failures; alternatively self-confidence has an executive function. It drags you out of bed, make you undertake things and gets you going. In the former case, self-esteem is a direct hedonic component of the utility function; in the latter case, it is not.

⁹ The attention paid to sunk costs in decision making is a much broader phenomenon. See, e.g., Dawes (1998).

To take another example, there is a lot of evidence that people refuse “unfair” offers. For example, in ultimatum game experiments, the “responder” often prefers to take 0 when the (anonymous) proposer offers to divide a cake of 100 into 90 for the proposer and 10 for the responder. Now, do you refuse an unfair offer in an ultimatum game because you like to punish others who are not nice to you (you are experiencing an emotion),¹⁰ or because you have a wired in (and emotionless) preference for fairness,¹¹ or else because you are just preoccupied with your self-image and you do not want to feel you are a weakling? The three possibilities¹² have distinct positive and normative implications.

Thus consider the consequences of (a) intention-based preferences, i.e., an intrinsic preference for punishing (being nice to) others who are not nice or intend not to be nice (are nice or intend to) to oneself, (b) social preferences, i.e., a wired-in preference for fair outcomes, and (c) a preoccupation with self image. The existence of a reciprocating action (e.g., the ability to punish the other) benefits the individual in cases (a) and (b) by enlarging the individual’s choice set; for example it may allow the unfairly treated individual to even the scales (not to mention to let off steam). In contrast, in case (c), the individual is induced to engage in costly self-signalling, and may well be better off when having no possibility of retaliation. On the positive side, (a) and (b) are not equivalent, as retaliation in case (a) may lead to unfair outcomes (the punishment need not be commensurate with the outrage). The different hypotheses also have quite different implications in terms of path dependency; for example, the past behavior of others plays a key role in case (a), while one’s past behavior is crucial in case (c).

So the bottom line is that we need to investigate the drivers of behaviors. A good argument can be made, though, that the development of rigorous experimental methods will reduce a bit the need for parsimony. By combining experimental and theoretical research and fostering a close interaction between the two, we will gain knowledge as to what concerns enter the utility function, and may feel somewhat less insecure about adding nonstandard hedonic components. A good example of such an approach is provided in Fehr and Schmidt (2001)’s survey, in which they use various standard games (ultimatum games, dictator games, gift exchange games, trust games, public good games) to systematically discuss the pros and cons of various theories for why a large fraction of people voluntarily forgo monetary gains in certain experiments, in particular theories of fairness based on the existence of social preferences or on intention-based reciprocity. I will not develop this further and refer the reader to Fehr and Rabin’s lectures for detailed discussions of preferences for fairness.

¹⁰ Rabin (1993) models such emotions using psychological game theory (Geanakoplos et al., 1989).

¹¹ As noted earlier, Fehr and Schmidt (1999) consider utility functions featuring inequity aversion. With two individuals, the utilities take the following form (in an atemporal context): $u^i(c^i, -|c^i - c^j|)$. Such preferences can rationalize both positive and negative actions towards other players.

¹² There is a large body of literature offering alternative explanations: See Fehr and Schmidt (2001) for a review.

2.2. Discounting: $\max_{A_t} E[\sum_{\tau \geq t} \delta^{\tau-t} u_{\tau}(c_{\tau}) | I_t]$

The second extensively questioned assumption is our modeling of discounting. A large literature, dating back at least to the ancient Greeks, has questioned the time consistency property that is built into the classical representation of homo economicus.¹³

There is a large amount of evidence showing that this time consistency assumption is only a rough approximation. We all have a taste for immediate gratification that sometimes makes us engage in behaviors that from an earlier, “colder” point of view we would like to avoid: we overeat or overdrink, we enter into fights we would rather avoid, we watch TV and delay unpleasant tasks at a high future cost.

These instances of time inconsistent behavior are the “tip of the iceberg” of the self-control problem. More interesting is the submerged part. To address our self-control problem, we often use external and internal commitments. External commitments include the purchase of season tickets to go to the opera with friends, the commitment of writing a paper for a conference, or the absence of television or cigarettes at home. Attempts at internal commitment include a variety of resolutions, targets and behavioral principles such as never drinking alcohol. The role of these resolutions and other cognitive gimmicks is to make impulsive behaviors (lapses) more salient and thereby more costly.

While classical economists have often discussed the self-control problem, it is only in the last decade that its consequences have been systematically investigated. Most of the recent work uses an unsatisfactory, but convenient and ultimately useful “hyperbolic” representation in which for instance, the current self (self t in the transparency) puts extra weight $1/\beta$ greater than 1 on his current costs and benefits $u_t(c_t)$:

$$\max_{A_t} E[u_t(c_t) + \beta \sum_{\tau \geq t+1} \delta^{\tau-t} u_{\tau}(c_{\tau}) | I_t].$$

Thus, self t will take decisions that are too short-termist from the point of view of previous selves, who do not suffer from the same instant gratification bias with respect to date- t decisions.

2.3. Beliefs: $\max_{A_t} E[\sum_{\tau \geq t} \delta^{\tau-t} u_{\tau}(c_{\tau}) | I_t]$

A third set of departures from homo economicus relate to the individuals’ belief formation or judgment under uncertainty.

Here again, there are too many bodies of evidence and citations to report. Let me just recall a few familiar ones. First, it will not come as a surprise to anyone that people make some systematic mistakes in updating probabilities. For example, some anomalies refer to inferences made by the individual concerning her environment on the

¹³ E.g., Smith (1759), James (1890), Böhm-Bawerk (1981), Ainslie (1992), Strotz (1956), Phelps and Pollak (1968), Laibson (1994, 1997) and Carrillo and Mariotti (2000). While hyperbolic preferences have gained much prominence in recent years, there have also been dissenting voices, and alternative ways of thinking about the self-control problem have been proposed (which for conciseness I will not discuss here): see in particular Caillaud et al. (1999), Gul and Pesendorfer (2001) and Krusell et al. (2001).

basis of a small number of observations. The gambler's fallacy refers to an excessive tendency people have to overinfer from short sequences and predict regression to the mean (for example, to overpredict the probability of "tails" after a sequence of four "heads"); the hot hand phenomenon in contrast refers to the overprediction of future events' correlation with past ones, as in the case of the exaggerated belief in streaks in sports. Much recent research has tried to reconcile these two observations reflecting biases toward negative and positive serial correlation, respectively.¹⁴

Psychologists have also documented a bias toward interpreting evidence as confirming one's prior beliefs as well as a hindsight bias ("I knew it all along").¹⁵ These two biases are self-serving in that they lead to an optimistic assessment of one's ability to predict. An interesting question is therefore whether such biases are related to another and well-documented bias regarding updating about one's self. An almost universal attitude is to attribute successes to oneself and failures to others or lack of luck (just think about how we react when receiving hostile referees reports!).¹⁶

Relatedly, there is a large amount of evidence that people's recollections of their past actions and performances are often self-serving: they tend to remember (be consciously aware of) their successes more than their failures, reframe their actions so as to see themselves as instrumental for good but not bad outcomes, and find ways of absolving themselves by attributing responsibility to others.

More generally, at each instant t the individual is aware of only a small subset of the information he has received in the past. Mathematically, the conditioning partition I_t refers to what comes to the individual's awareness at the moment of decision making and is usually much coarser than the partition corresponding to the full set of signals received by the individual's up to date t . And, quite interestingly, this imperfect recall is *endogenous*, as it depends on strategies, such as emotion and attention control that we may employ when we receive and encode information, as well as other strategies, such as cue management and rehearsal, that will make it more or less likely that this information be present in awareness and thereby used in future decisions. Our knowledge of the functioning of memory is still very limited, but psychologists have documented some patterns that, in my view, we have made insufficient use of.¹⁷

A last category of judgment biases, called projection biases, relates to the tendency to look at others (other people or future selves) from the point of view of one's current self.¹⁸ For example, Gilbert et al. (1999) have documented that affective forecasts (think of this as the forecast of $u_\tau(c_\tau)$ by self t for τ greater than t) have a durability bias. People who are asked about the hedonic consequences that would result from their learning that they have a cancer, that they are denied tenure, or that their marriage is about to break up underestimate their psychological immune system;

¹⁴ Mullainathan (2001), Rabin (2000a) and Rabin and Vayanos (2000).

¹⁵ E.g., Tversky and Kahneman (1971) and Gilovich (1991).

¹⁶ E.g., Baumeister (1998) and Festinger (1954).

¹⁷ E.g., Anderson (2000). See e.g., Laibson (2001), Mullainathan (1998) and Piccione and Rubinstein (1997) for some economic implications of imperfect recall.

¹⁸ For some implications of projection biases, see Loewenstein et al. (1999).

that is, their affective state will return to baseline much faster than they predict.¹⁹ More generally, people have a hard time making affective forecasts, and experienced and decision utilities may differ.²⁰

2.4. Optimization: $\max_{A_t} E[\sum_{\tau \geq t} \delta^{\tau-t} u_{\tau}(c_{\tau}) | I_t]$

Last, psychologists often question our discipline's maximization hypothesis. We do make lots of mistakes, although I am not sure that these mistakes should necessarily be viewed as instances of irrational behaviors. Quite plainly, thinking through the analytics of decision making (for me at least) is time consuming and costly. This observation has simple implications such as the incomplete self-description of the problem at hand, the limited use of backward induction,²¹ the cost to the decision-maker of adding options in A_t ,²² and most importantly, the heuristics that we use to solve problems. By their very nature these heuristics are tremendously helpful on average even though they are very poorly adapted to certain types of situations.

In this spirit, when I see experimental evidence such as that underlying the Allais paradox, I find it very unproductive to marvel at the stupidity or irrationality of the human mind (this reaction may be attributed to the fact that I almost invariably fall into the traps built by the experimenters!). I consider it much more productive to use the Allais paradox and the like to try to retrieve information about the rules of thumb, mental representations, analogies and categorization strategies that we employ. These heuristics are probably quite efficient "on average" even though they may be very inappropriate in specific decision-making environments.²³ I would make very similar remarks regarding the (very pervasive) framing effects, which refer to the ways in which perceptions and choices may depend on the formulation of the problem; for example cash discounts seem to be perceived differently by consumers from card surcharges or reward schemes seem to be viewed by workers as differing from punishment schemes.

A type of evidence that is potentially more damaging for the rationality hypothesis comes from self-destructive behaviors. Examples of such behaviors that we will later discuss are self handicapping – taking actions that impair one's performance: underpreparing or drinking before an exam, withholding of effort, self-setting of over-ambitious goals or tasks –, and the flip side of impulsive behaviors, compulsive ones, the behaviors of people such as workaholics and anorexics who apply constant pressure to themselves and suffer from "overcontrol".

¹⁹ One way of categorizing the findings of Gilbert et al. as a projection bias is to view the current self as "thinking" that the bad news will be as vivid and present in the future selves' awareness as they are in the current self's awareness.

²⁰ E.g., Kahneman et al. (1997).

²¹ E.g., Gabaix and Laibson (2000).

²² E.g., Mirrlees (2000).

²³ There is a large literature on heuristics, e.g., Gigerenzer et al. (1999), Gigerenzer (2000), Kahneman et al. (1982), Mullainathan (2001), Plott and Smith (2000) and Tversky and Kahneman (1974).

3. How should we respond to the challenge?

A first possible reaction is,

so what? These phenomena are interesting but best left to psychologists, as they may end up being second-order effects for economic matters.

It is too early to tell, but a behavioral view will probably prove helpful in some areas of economics. For example, a number of topics such as framing and self-image are bound to bear on our views concerning advertising and marketing. Other fields that seem ready for cross-fertilization are organizational behavior, sociological economics, health economics, and finance.²⁴

A sizeable fraction of our profession, let us call them “the rationalists”, is skeptical about the scope for cross-fertilization. Rationalists are not ignorant of the shortcomings of the homo economicus paradigm. Rather, and oversimplifying, they are concerned that the behavioral approach will prove too impulsive for its own sake and may forget what economics is all about, namely parsimony and normative analysis. They also point out that the experiments conducted by psychologists and that motivate much of the new literature are not proper tests of formal theories and that one needs to integrate experimental research with theoretical research in a more structural approach.

We face a trade-off that is familiar and common to all attempts to enrich and increase the realism of the existing paradigm. The profession then is torn between the fast-return strategy of taking shortcuts and using reduced forms exposed to critiques such as the Lucas critique, and the longer-term and more uncertain strategy of building foundations on a small number of reasonable hypotheses.

“Reduced forms” and “shortcuts” should not necessarily be interpreted as derogatory statements. After all, the Arrow–Debreu general equilibrium theory – a cornerstone of our discipline – has been built on the (crazy) premise that prices were quoted by a Walrasian auctioneer, with the foundations coming much later.²⁵

Keynesian theory is another and perhaps more interesting example of pragmatism that paid off. By every standard, this theory was sloppy and full of uncertain shortcuts. Only recently have we started to sort out those predictions that are robust to better modeling from those that are not; but one must bear in mind that many of the building blocks required to lay proper foundations (imperfect competition, search theory, corporate finance, dynamic analysis, etc.) were developed 40–50 years after the General Theory and some are still to be developed! In the meantime, ISLM helped us think about policy and created a body of hypotheses and facts that the new foundations could be confronted with.

But pragmatism does not always pay off. Part of the profession in the 50s and 60s revolted against the view that firms obeyed the neoclassical rationality embodied in profit maximization. Rightly so: firms often do not seem to maximize profits! A large

²⁴ A field that has already received much attention, including for example some remarkable work by David Laibson, by himself (e.g., Laibson, 1994, 1997) and with coauthors, e.g., Harris and Laibson (1999), and by Mullainathan (1998).

²⁵ To be fair, general equilibrium theorists had little worry that foundations in terms of large-number oligopolistic competition would later come about.

literature involving some of the brightest people of the profession then developed, that included in the firm's objective function not only profit, but also revenue, sales, size of the staff, growth rate, employment stability, and any other variable that seemed relevant.²⁶ If you go back and read this literature, these hedonic additions to the firm's objective function were actually well motivated.²⁷ But this line of research eventually did not abide by its promises, and soon thereafter was displaced by a new theory with more explicit foundations – agency theory –, that later proved extraordinarily fruitful. As usual, the reduced forms presumed by the managerial-objectives-of-the-firm literature were not stable to changes in the firm's environment. For example, the weights put by managers on firm size or sales depend, as we know, on corporate governance. So, for example, a change in the legal environment (say, a new regulation affecting the mechanics of corporate control) affects the managers' ability to build empires or to select investments so as to entrench themselves.

I am reluctant to take a position on the future of psychology and economics. We are still in the exploratory phase and any stance runs the risk of pushing the field in the wrong direction. It is important, even crucial, to “let a thousand flowers bloom”; there will be plenty of time for the shake-out period later on. My own and evolving preference is to attempt to enrich *homo economicus* by embodying a small number of new ingredients. My gut feeling – or maybe it is just wishful thinking – is that many of the phenomena that I discussed earlier have a small number of common foundations and that we may be able to benefit from the evidence on human behavior assembled by psychologists without necessarily losing the virtues of economic analysis – parsimony and welfare analysis.

This gut feeling has guided my own research with Roland Benabou in a particular area of psychology and economics, research on which I would now like to report.

4. Self-confidence and personal motivation

This research is concerned with the role of self-perception and self-image in social and economic interactions, or “*egonomics*”.²⁸ Its goal is to develop a simple economic model building on and unifying a wide variety of themes in psychology in order to explain a range of apparently “irrational” behaviors documented by psychologists: unwillingness to know, self-handicapping, distorted attention, selective memory and self-deception, compulsive behaviors, as well as a number of organizational/sociological phenomena, such as the possibility that contingent rewards undermine intrinsic motivation, that help may reduce autonomy, that people may both boost and bash the others' ego, and that people may self-deprecate.

²⁶ E.g., Baumol (1959), Leibenstein (1960), Marris (1965) and Williamson (1964).

²⁷ Indeed, it would be futile to use the benefit of hindsight to conclude that it should have been obvious from the start that the contributors to this literature were just mistaken. Not only were many of the contributors innovative and respected scholars, but their motivations for altering the classical paradigm were also refined and grounded in observation. The managerial-objectives-of-the-firm literature anticipated many of the themes of the subsequent agency literature (see e.g., Baumol, 1959, Chapter 6).

²⁸ To use a word coined by Schelling (1978).

4.1. *The demand for self-confidence*

Let us begin with the demand for self-confidence. In most societies, self-confidence is widely regarded as a valuable individual asset. Today an enormous “self-help” industry flourishes and American schools and companies place a strong emphasis on imparting children and employees with self-confidence. Why is a positive view of oneself, rather than a fully accurate one, so desirable?

A first reason emphasized by psychologists is that thinking of oneself favorably may just make a person happier. Self-image θ is then simply an argument in the utility function, $u_\tau(\theta)$. While it is obviously “part of the story”, a theory of self-confidence based on these affective benefits however confronts some difficult questions. First, it may face an embarrassment of riches with regards to the personal traits θ that people value on a per se basis. Second, even if the trait θ stands for some usually desirable attribute such as “ability”, the monotonicity assumption may not be a foregone conclusion. While people usually like to think of themselves as talented (want to enjoy blissful optimism), the same people also occasionally engage in defensive pessimism, that is attempt at minimizing previous achievements. Along similar lines, it is not a priori clear that the function $u_\tau(\theta)$ should be assumed concave or convex, even though this choice has important implications for attitudes toward information acquisition.

Last, a purely affective theory of self-confidence does not readily extend to social and economic interactions, where people clearly seek out optimistic, self-confident partners rather than depressed, self-doubting ones. By contrast, the instrumental approach that I will shortly develop offers a unified explanation of why and when self-confidence is valuable for ourselves and for others with whom one interacts. This instrumental approach is based on the idea that a more accurate self-knowledge has costs and benefits for decision making. On the benefit side, knowing who you are means knowing your limits, that is avoiding overconfidence. On the cost side, thinking positively may preserve your motivation.

4.2. *Grains of sand: $\max_A, E[\sum_\tau \geq t \delta^{\tau-t} u_\tau(c_\tau) | I_t]$*

We introduce three grains of sand in the homo economicus paradigm: imperfect self-knowledge, imperfect willpower, and imperfect recall (for applications to social interactions only the first is needed). Imperfect self-knowledge is hardly a grain of sand. But although economic theory sometimes allows for imperfect self-knowledge,²⁹ homo economicus usually knows who he is; psychology in contrast stresses the process of learning about oneself and the struggle with one’s identity. We thus assume that the individual is uncertain about parameters such as his talent or his ability to resist impulses.

Next, to give content to internal conflicts and self-defeating behavior, we allow for imperfect willpower in the form of hyperbolic-like discounting. Time inconsistency

²⁹ For example, individuals may not know their taste for a particular product (e.g., Farrell, 1986), their talent (e.g., Holmström, 1999), or their propensity to addiction (e.g., Orphanides and Zervos, 1995).

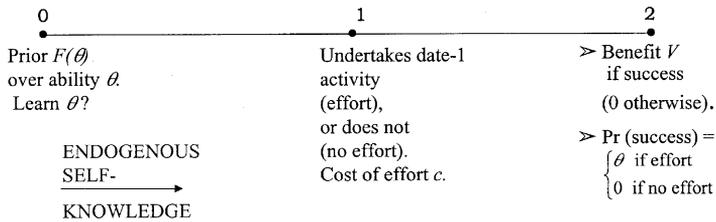


Fig. 1.

combined with imperfect self-knowledge will create scope for rational ignorance, as in Carrillo and Mariotti’s seminal work.

Last, to account for motivated cognition and self-serving beliefs, we allow for (exogenously determined) rates of recall. We give ourselves few instruments, though, as we keep the assumptions that selves maximize their intertemporal utilities, and that they employ Bayes rule (or more generally are not too naive).

4.3. *The value of information*

Our work builds on the work of Carrillo and Mariotti (2000), who have the first shown that time inconsistency creates scope for rational ignorance.

To take a simple example, summarized in Fig. 1, consider a risk neutral individual with a three-period horizon, $t = 0, 1, 2$. At date 1, this individual will decide whether to exert effort (undertake a task) at positive cost c or not. Effort results in probability of success θ at date 2, where θ can be interpreted as “ability” in the task. Success brings about benefit V to the individual, and failure yields 0. The individual knows all parameters except θ ; the prior over this ability parameter is denoted by $F(\theta)$. [We assume imperfect knowledge of θ , but it could equivalently be imperfect knowledge of cost c or payoff V].

Consider now the following thought experiment: Someone at date 0 offers to reveal the true value of θ to the individual for free. Will the individual accept to learn the truth? Clearly yes if the individual has time consistent preferences, as we know that the value of information is then always positive.

Not so if the individual knows that he will face a self-control problem in the future. Self 0 may be concerned that new information might encourage self 1 to procrastinate.

To see this, let us model time inconsistency in the familiar way:

Self 1’s payoff is

$$u_1 + \beta\delta u_2 = -c + \beta\delta\theta V \quad \text{when exerting effort,}$$

$$0 \quad \text{otherwise.}$$

Self 0’s payoff is

$$u_0 + \beta(\delta u_1 + \delta^2 u_2) = u_0 + \beta\delta(-c + \delta\theta V) \quad \text{when self 1 exerts effort,}$$

$$0 \quad \text{otherwise.}$$

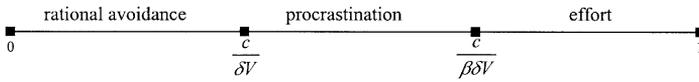


Fig. 2.

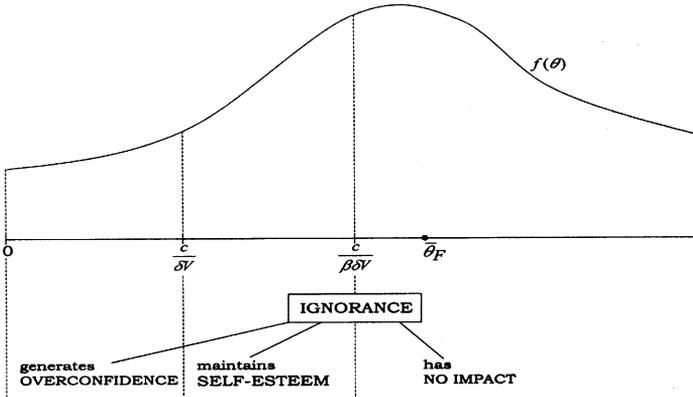


Fig. 3.

When deciding whether to exert effort, self 1, who has a taste for instant gratification, perceives the current cost to be c/β with $\beta < 1$, to be compared with discounted expected benefits, $\delta\theta V$, from effort (δ is the standard discount factor). In contrast, self 0 faces no instant gratification bias with respect to future payoffs and compares $\delta\theta V$ to c rather than to c/β . In the case in which the true value of θ is known to the individual, we are thus led to consider three regions: See Fig. 2. For high self-confidence (θ exceeds $c/\beta\delta V$), self 1 exerts effort; there is no self-control issue. For low self-confidence (θ lower than $c/\delta V$), selves 0 and 1 both benefit from the task not being undertaken; again, self-control is not an issue as there is no dissonance between the two selves' preferences. In contrast, for intermediate levels of self-confidence, self 1 procrastinates, that is does not exert effort even though from the point of view of date 0 he would be better off if he could commit to exert effort.

We can now derive the value of information.

Suppose that the individual starts with strong self-confidence as in Fig. 3. The prior distribution has mean $\bar{\theta}_F$ high enough to induce effort in the absence of new information. Remaining ignorant rather than learning the true θ has no impact if the true ability lies in the right-hand region ($\theta > c/\beta\delta V$). Ignorance is costly if the true ability turns out to be very low (in the left-hand region, $\theta < c/\delta V$), since ignorance leads the individual to undertake a project he is no good at, an instance of overconfidence. The individual would be better off knowing his limits. In contrast, ignorance is a good thing if the true θ lies in the intermediate region, since it then prevents procrastination by avoiding damages to the individual's self-esteem. Intuitively, the value of information may be negative except when β is close to 1 and so the intermediate region is almost empty (that is, when the individual's preferences are almost time consistent).

We can exploit this simple example a bit further. What kind of individual is more likely to not want to receive ego-related information? Comparing initial self-confidence levels really involves a comparison of distributions. A natural comparison goes as follows: An individual with distribution F over ability θ has higher self-confidence than an otherwise identical individual with distribution G if the likelihood ratio is monotonic, i.e., if the ratio of the densities $f(\theta)/g(\theta)$ is increasing in θ .

It is straightforward to show that people with higher initial perceptions of their abilities are also the more insecure about receiving potentially ego-threatening information, as they are relatively more concerned that bad news will make them fall into the procrastination region, and relatively less concerned that their talent is very low (in the overconfidence region). More formally, if an individual prefers not to receive information in order to preserve his self-confidence, so will any individual with higher initial self-confidence. Intuitively, and referring to Fig. 3, an individual with a higher self-confidence has a higher relative probability of falling into the intermediate (procrastination) region than into the left-hand (overconfidence) region.

Rational ignorance is closely related to *self-handicapping*. Recall that self-handicapping refers to behaviors such as inadequate preparation, withholding of effort, drinking before a task, or the choice of undoable tasks. Such strategies can be viewed as a date-0 choice by the agent to make his date-0 performance less informative about his ability. For example, an undoable task never results in bad news about oneself. Should we then conclude from our previous analysis that people with higher self-confidence are more likely to self-handicap? We indeed just saw that individuals with higher self-confidence are more at risk when confronted with self-relevant information. There exists a countervailing effect, though: In many situations, an individual with a higher self-esteem also incurs a higher opportunity cost when he chooses to self-handicap in order to protect his self-esteem. That is, he is probably good at the date-0 task and so he foregoes a higher expected return when self-handicapping. So we should not expect sharp predictions concerning the relationship between self-confidence and self-handicapping; the empirical evidence in this respect is no more conclusive than the theory.

Rather than initial self-esteem, the personal trait which the model suggests should be robustly correlated with a subject's propensity to engage in willful ignorance, self-handicapping and the like is their degree of undermotivation, that is, their (self-perceived) bias towards instant gratification. The economic approach to self-regulation – common to our work and that of Carrillo and Mariotti (2000) and Brocas and Carrillo (1999) – thus suggests a link between two hitherto disjoint areas of experimental psychology, namely those on intertemporal preferences and self-esteem maintenance.

5. The psychological immune system

We have until now focused on the decision of whether to acquire self-relevant information. A fair amount of belief manipulation, though, occurs through the management of feedback about our performance and ability from parents, teachers, spouses,

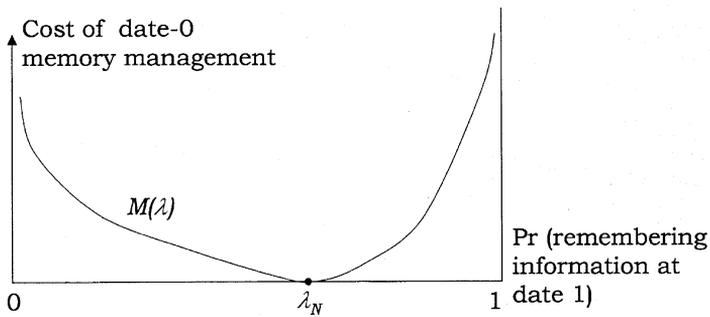


Fig. 4.

coworkers or simply from observing our performance and comparing it with that of others.

Psychologists, philosophers and writers have long documented people's universal tendency to rehearse good news and to deny, explain away and selectively forget ego-threatening information. At the same time, they have pointed out that we cannot just choose our beliefs (would not it be nice if we could?) and that this impossibility of choosing beliefs we like stood in the way of a fully consistent theory of self-deception. For instance, as Sartre (1953) pointed out, how can the individual simultaneously know and not know the same information? How can we relate hot attributions and motivated beliefs, which are universally viewed as self-esteem maintenance strategies, and cold attributions, resulting (as economic theory presumes) from a rational cognitive process?

Our approach to reconciling the motivation and cognition aspects consists in unbundling the "self that knows" from the "self that doesn't know" by using standard observations about the imperfection of memory.

The basic idea is that the individual can, within limits and at a cost, affect the probability of remembering a given piece of information. Recall that people (consciously or unconsciously) try to remember good news (self-enhancing information) by paying more attention when they receive them and then by rehearsing them (since rehearsal, as we know, increases the probability of remembering – this is the reason why we cram for an exam): they discuss these good news, keep hard evidence, leave it on their desk, and tend to prefer and therefore select environments and interactions that remind them of ego-favorable information. Conversely, when receiving self-threatening information, people tend to pay limited attention and to create distractions (for example fight). They seek contradictory evidence and excuses ("the referee is biased or incompetent", "I was tired when I gave that seminar in which I was criticized") and rehearse them. And they avoid negative cues later on.

A simple way of capturing such strategies (and one consistent with a more fundamentalist approach) is to assume that when receiving information at date 0, the individual can, at a cost, increase or decrease the probability that this information be brought back to awareness at date 1, like in Fig. 4. λ_N is a "natural rate of recall"; in some instances this rate of recall may be high and hard to bring down, as when one becomes obsessed with past failures or when one cannot sleep because there is

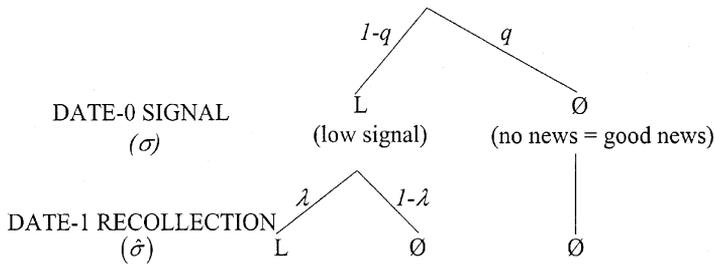


Fig. 5.

an important exam or meeting on the morrow. But the rate of recall is in general manipulable.

To make things more concrete, consider the date-1 procrastination problem discussed in Section 4 and suppose that at date 0, the individual may or may not receive a self-threatening signal about his ability θ : see Fig. 5.

Thus the date-0 signal is with some probability a low signal, L . Receiving no signal (denoted by the empty set symbol) is good news. The individual cannot recollect signals he did not receive, but he will with endogenous probability λ recall the bad signal if he receives it. With probability $1 - \lambda$, self 1 will not recall the signal and therefore will be confused as “no recollection” may mean either good news or bad news that he is no longer aware of. A Bayesian self 1 discounts the absence of recollection of a bad signal as he knows that he has an imperfect memory. Note also the externality: self 1 discounts the good news all the more that self 0 has an incentive to repress the bad news, that is to choose strategies that lower λ . This self-doubt is the individual’s Bayesian reaction to his self-serving beliefs. Just like a ruler whose entourage dares not bring him bad news, or a child whose parents praise him indiscriminately, an individual with some understanding of the self-serving tendency in his attention or memory can never be sure that he really “did great”, even in instances where this was actually true.

This externality across states of nature is the new feature introduced by imperfect recall. The manipulation of information by self 0 when a low signal accrues (i.e. the choice of λ) otherwise exhibits the familiar tradeoff between maintaining self-esteem to prevent procrastination (which calls for forgetting – λ low) and avoiding overconfidence (a risk that is reduced if one remembers one’s limits, i.e., if λ increases).

Last, we assume that each self optimizes given his available information; that is we focus on perfect Bayesian equilibria.

The typical outcome is as in Fig. 6.

First, the individual engages in selective memory, and more so when he is more time inconsistent (has a lower β). Indeed the benefits of confidence-building rise with time inconsistency, while the risk of overconfidence decreases with it. Second, for intermediate degrees of time inconsistency, there are multiple equilibria. The individual may engage little in repression (λ is high); the individual then tends to take recollections at face value, which in turn increases the risk of overconfidence in case of repression

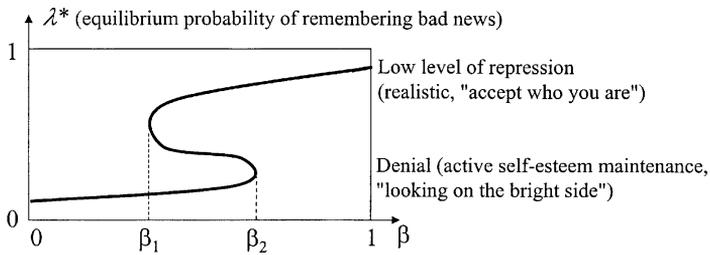


Fig. 6.

and thus encourages him to accept who he is. Or he may engage in denial and actively maintain his self-esteem.

Another insight of a theory of memory management is that because of the spoiling of good news by self doubt, the individual may be caught in a self-trap, defined as an inferior equilibrium in which all selves in the strong form, or at least self 0 in the weak form would like to move to a superior equilibrium, or, if the equilibrium is unique, would be better off if the recollection rate λ could be constrained through the use of external devices. To put things in perspective, recall that there is an enormous industry of “self-help” books, courses, gurus and now web sites claiming to help people improve their self-esteem and that of their children. But is a person ultimately better off following a strategy of active self-esteem maintenance and “positive thinking”, or when he always faces the truth? Psychologists appear sharply divided between these two conflicting views of self-deception. On one side are those who endorse and actively promote the self-efficacy/self-esteem movement, pointing to studies which tend to show that a moderate dose of “positive illusions” has significant affective and functional benefits. On the other side are skeptics and outright critics who see instead a lack of convincing evidence, and point to the dangers of overconfidence as well as the loss of standards which results when negative feedback is systematically withheld or discounted in the name of self-esteem preservation.

Our formal analysis helps provide insight into the reasons for this ambiguity. A selective memory is good when self-confidence has been hurt by bad news as the individual optimally trades off self-esteem maintenance with the risk of overconfidence, but hurts in a high state due to the self-doubt effect. Furthermore, self-deception is more likely to be beneficial (as well as more likely) for a less time-consistent individual.

Motivated memory management also underlies *defensive pessimism*, the individual’s occasional attempts to minimize his previous achievements or to convince himself of the difficulty of the future task.

We have made the reasonable assumption that ability and effort are complements in generating future payoffs: trying is worth more if one is talented. In some situations, though, ability and effort are substitutes. For example, cramming for a pass–fail exam makes less of a difference for a good student who knows he will probably pass anyway than for an average student. Worrying about procrastination before the exam, the good student may want to attribute his past grades to luck or the easiness of past exams

in order to avoid coasting or slacking off. Defensive pessimism is the counterpart of confidence enhancement when ability and effort are substitutes rather than complements. Thus the same theory predicts blissful optimism and defensive pessimism, depending on the nature of the production function.

6. Willpower and personal rules

The cognitive strategies discussed so far fall short of capturing what is usually meant by self-regulation, namely the various attempts at responding to one's self-control problem. I mentioned earlier the external commitment devices, which can be viewed as taking date-0 actions that change the actual values of the ratio $\theta V/c$ rather than altering the beliefs about it. For example, committing to giving a paper at a conference increases the stake associated with not having a satisfactory version by the date of the conference. All of us also use internal commitment devices. We adopt and rehearse behavioral principles: follow a diet, exercise every other day, drink at most one glass of wine at diner, smoke only one cigarette after each meal, always finish what one has started, do onto others... .

To understand the rationale for such rules, we propose to start from psychologists' insight that people often have imperfect knowledge of their willpower either overall or in a specific activity. They may worry that their β may not be sufficiently high or the cost c sufficiently low when they experience craving, and so that they may not have enough willpower to persevere in whatever activity they attempt. Ainslie (1992), Baumeister et al. (1994) and others have argued that this form of imperfect self-knowledge creates a concern for self-reputation as lapses are treated as precedents and discourage future attempts at reaching high personal payoffs. Note that this view requires imperfect memory. The importance given to a past lapse, that is actions taken by the individual, means that the individual has only an imperfect recollection of the intensity of craving that led to the lapse. This suggests looking at the following setup depicted in Fig. 7.

Fig. 7 depicts a stage game of a repeated situation. The stage game is decomposed into two substages: substage I in which the individual "tries" or "does not try". Trying for example means "undertaking an ambitious research project" or "not consuming cigarettes, alcohol or food". As the temptation becomes more intense, we get to a "craving substage", substage II, in which the individual must choose whether to pursue

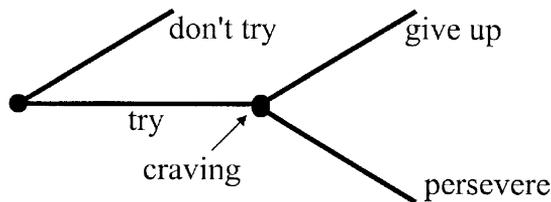


Fig. 7.

the hard project in the face of serious difficulties, or to keep staying away from cigarette, alcohol or food.

We make three assumptions: (1) the individual has imperfect information about his ability to resist craving and persevere (that is, he has imperfect information about the β that will prevail at substage II or about the cost c of persevering); (2) the individual is too tempted not to try in the first place (this part is just the standard self-control story – not trying provides instant gratification); (3) partly because of the difficulty of codifying levels of craving and partly because of belief manipulation of the type discussed earlier, the individual cannot recall precisely the intensity of craving from past similar experiences and therefore in future stage games will use past behavior (did I give up on the research project, or resume smoking cigarettes?) to obtain information about this intensity of craving. This third assumption underlies the self-monitoring of actions, i.e., the importance given by the individual to lapses.

From this model, we derive the following predictions: The individual is more likely to try (that is his willpower is more likely to be put to the test) and he will exert more restraint in the following circumstances:

- First, if the situation is repeated. This conclusion is not surprising in a self-reputation model, but it has some interesting consequences. In particular, it can be shown that early externally forced behavior – that is, when perseverance is forced by the parents, social norms or various extrinsic incentives rather than the object of individual choice – reduces future autonomy (the individual is less likely to try in the future).
- Second, the individual is more likely to try and exert self-control if lapses are less forgettable. Again, no surprise in a self-reputation context and yet interesting implications, in particular the rationalization of religions, morals, behavioral principles, resolutions, and self-setting of targets as cognitive ways of making sure that lapses are particularly salient and so the individual is concerned about acting impulsively and thereby violating the rule.
- Third, the existence of self-excuses (technically, nonpersistent shocks to the cost of persevering – introducing noise) reduces self-restraint. Self-excuses allow the individual to find a plausible external motivation for what is really a lapse.

To conclude this section, I want to stress an important point. While attempts at self-reputation and self-restraint generally benefit the individual as they help solve his self-control problem, they may “overshoot” and lead him into a self-trap. That is, self-restraint may be excessive. We have all seen people jogging in the rain or the snow. More serious cases involve workaholism, anorexia, miserliness or oversaving (old people keeping the same saving rate even though they have no heirs), and so forth. Compulsiveness is the flip side of impulsiveness. It is precisely because people are concerned about possible impulsive behaviors (engaging in procrastination, overeating, or being a spendthrift) that they develop compulsions. Economists often criticize hyperbolic discounting on the ground that people sometimes seem to exhibit a salience for the future rather than one for the present. Our point is simply that compulsiveness arises precisely as a response to a perceived self-control problem. An inefficient response but an equilibrium one!

7. Concluding remarks

Let me conclude this lecture with two points.

The first is that one of the most fruitful alleys of research for economists is to embody behavioral models in their social context. As we argued, a unified approach to social psychology should start from a *single* view of the individual's preferences, cognitive machinery and basic problem-solving strategies. While incentives and feedback, and therefore behavior, are highly context-dependent, the underlying "fundamentals" are the same whether the individual is engaged in self-regulation or interacting with others. For example, Roland Benabou and I have looked at the management – for example, by parents and coworkers – of other people's self-esteem, assuming that the latter have imperfect self-knowledge. Reporting on this research is out of the scope of the lecture,³⁰ but we think that it can help refine our views on organizational behavior. For example, we can study how and when providing people with contingent rewards (incentives) can backfire and reduce intrinsic motivation. In particular, we relate perverse effects of contingent rewards to the existence of private information held by the reward provider about the agent's ability, difficulty of the task, or eventual payoff.

With Marco Battaglini,³¹ we have studied why people may seek relief from their self-control problem by belonging to peer groups with specific characteristics such as Alcoholic Anonymous, and more generally when and why social interactions may alleviate or aggravate the individual's self-control problem.

By and large, this area of social psychology is territory untrodden by economists and seems extremely promising. Topics include status relationships, envy, and community identity and relationships, but the reader will easily figure out several others.

Last, and to return to a previous theme, I feel that we (I include myself, since I suffer from self-control problems) ought to focus less on outcome – the behavior – and to go deeper into its foundations than economists, and to some extent psychologists do; that is, I think we should go "pico" where we currently go "micro". On the question of affect, we need to design experiments that distinguish better among the various possible motivations for documented behaviors. On the cognitive side, progress in understanding the way we categorize matters and develop heuristics and in understanding memory and awareness seems particularly important to understand departures both from Bayes rule and from full-fledged maximization. On both the affect and the cognitive sides, I feel that we should aim at building our theories on parsimonious foundations whenever possible.

I wish I were able to provide the reader with a clearer message as to where this new field is or should be going. It is probably in the nature of things that I am unable to do so. But I hope that this paper and those of Ernst Fehr and Matt Rabin in this volume will give a taste of what is going on and perhaps encourage some to join us in this difficult but exciting undertaking.

³⁰ Benabou and Tirole (2000).

³¹ Battaglini et al. (2001).

References

- Ainslie, G., 1992. *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person (Studies in Rationality and Social Change)*. Cambridge University Press, Cambridge, England and New York.
- Akerlof, G., 1984. Gift exchange and efficiency-wage theory: Four views. *American Economic Review* 74, 79–83.
- Akerlof, G., Dickens, W., 1982. The economic consequences of cognitive dissonance. *American Economic Review* 72, 307–319.
- Anderson, J., 2000. *Cognitive Psychology and its Implications*, 5th Edition. Worth Publishers, New York.
- Battaglini, M., Benabou, R., Tirole, J., 2001. Self-control in peer groups. Mimeo., Princeton University and IDEI, Toulouse.
- Baumeister, R., 1998. The self. In: Gilbert, D., Fiske, S., Lindzey, G. (Eds.), *The Handbook of Social Psychology*. McGraw-Hill, Boston.
- Baumeister, R., Heatherton, T., Tice, D., 1994. *Losing Control: How and Why People Fail at Self-Regulation*. Academic Press, San Diego, CA.
- Baumol, W.J., 1959. *Business Behavior, Value and Growth*. Harcourt, Brace and World, New York.
- Becker, G., Murphy, K., 1988. A theory of rational addiction. *Journal of Political Economy* 96, 675–700.
- Benabou, R., Tirole, J., 2000. Intrinsic and extrinsic motivation. Mimeo., Princeton University and IDEI, Toulouse.
- Bentham, J., 1970. *Introduction to the Principles and Morals of Legislation (1789)*. Athlone Press, London.
- Böhm-Bawerk, E., 1981. *The Positive Theory of Capital* (translated by W. Smart). Macmillan, London.
- Brocas, I., Carrillo, J., 1999. Entry mistakes, entrepreneurial boldness and optimism. Mimeo., ULB-ECARES, June.
- Caillaud, B., Cohen, D., Jullien, B., 1999. Towards a theory of self-restraint. Mimeo., CERAS, December.
- Caplin, A., Leahy, J., 2001. Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of Economics* 116, 55–80.
- Carrillo, J., Mariotti, T., 2000. Strategic ignorance as a self-disciplining device. *Review of Economic Studies* 67, 529–544.
- Constantinides, G., 1990. Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98, 519–543.
- Dawes, R., 1998. Behavioral decision making and judgment. In: Gilbert, D., Fiske, S., Lindzey, G. (Eds.), *Handbook of Social Psychology*, 4th Edition. McGraw-Hill, New York (Chapter 11).
- Farrell, J., 1986. Moral hazard as an entry barrier. *Rand Journal of Economics* 17, 440–449.
- Fehr, E., 2002. Incentives anomalies. *European Economic Review* 46, this issue.
- Fehr, E., Schmidt, K., 1999. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114, 769–816.
- Fehr, E., Schmidt, K., 2001. Theories of Fairness and Reciprocity. Evidence and Economic Applications. DP 2703. CERP, London.
- Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics* 108, 437–459.
- Fehr, E., Kirchsteiger, G., Riedl, A., 1998. Gift exchange and reciprocity in competitive experimental markets. *European Economic Review* 42, 1–34.
- Festinger, L., 1954. A theory of social comparison processes. *Human Relations* 7, 117–140.
- Frank, R., 1988. *Passions within Reason: The Strategic Role of Emotions*. Norton, New York.
- Frank, R., 2001. *What Price the Moral High Ground?* Princeton University Press, Princeton, NJ, in preparation.
- Gabaix, X., Laibson, D., 2000. Bounded rationality and directed cognition. Mimeo., Harvard University.
- Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. *Games and Economic Behavior* 1, 60–79.
- Gigerenzer, G., 2000. *Insightout: Rethinking Rationality*. Oxford University Press, Oxford.
- Gigerenzer, G., Todd, P., the ABC Research Group, 1999. *Simple Heuristics that Make us Smart*. Oxford University Press, Oxford.

- Gilbert, D., Pinel, E., Wilson, T., Blumberg, S., Wheatly, T., 1999. Immune neglect: A source of durability bias in affective forecasting. Mimeo.
- Gilovich, T., 1991. *How We Know What Isn't So*. Free Press, New York.
- Gul, F., Pesendorfer, W., 2001. Temptation and self-control. *Econometrica* 69, 1403–1435.
- Harris, C., Laibson, D., 1999. Dynamic choices of hyperbolic consumers. Mimeo.
- Holmström, B., 1999. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies* 66, 169–182.
- James, W., 1890. *The Principles of Psychology*. World Publishing, Cleveland, OH.
- Jevons, W., 1905. *Essays on Economics*. Macmillan, London.
- Kahneman, D., Knetsch, J.L., Thaler, R.H., 1990. Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy* 98, 1325–1348.
- Kahneman, D., Wakker, P.P., Sarin, R., 1997. Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics* 112, 375–406.
- Kahneman, D., Slovic, P., Tversky, A., 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge.
- Köszegi, B., 1999. Self-image and economic behavior. Mimeo., MIT, Cambridge, MA, October.
- Krusell, P., Kuruscu, B., Smith, A., 2001. Time orientation and asset prices. Mimeo.
- Laibson, D., 1994. *Essays in hyperbolic discounting*. Ph.D. Thesis, MIT, Cambridge, MA.
- Laibson, D., 1997. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 112, 443–478.
- Laibson, D., 2001. A cue-theory of consumption. *Quarterly Journal of Economics* 116, 81–120.
- Leibenstein, H., 1960. *Economic Theory and Organizational Analysis*. Harper & Row, New York.
- Loewenstein, G., 1987. Anticipation and the valuation of delayed consumption. *Economic Journal* 97, 666–684.
- Loewenstein, G., O'Donoghue, T., Rabin, M., 1999. Projection bias in predicting future utility. Mimeo.
- Marris, R., 1965. *The Economic Theory of Managerial Capitalism*. The Free Press, Glencoe.
- Marshall, A., 1881. *Principles of Economics*. Macmillan, London.
- Mirrlees, J., 2000. The economics of imperfectly rational agents. Presidential Address, European Economic Association, Delivered in Bolzano on August 31.
- Mullainathan, S., 1998. A memory based model of bounded rationality. Mimeo., MIT, Cambridge, MA.
- Mullainathan, S., 2001. Thinking through categories. Mimeo., MIT, Cambridge, MA.
- Orphanides, A., Zervos, D., 1995. Rational addiction with learning and regret. *Journal of Political Economy* 103, 739–758.
- Phelps, E., Pollak, R., 1968. On second-best national savings and game equilibrium growth. *Review of Economic Studies* 35, 185–199.
- Piccione, M., Rubinstein, A., 1997. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior* 20, 3–24.
- Plott, C., Smith, V., 2000. *Handbook of Experimental Economics Results*. North-Holland, Amsterdam.
- Rabin, M., 1993. Incorporating fairness into game theory and economics. *American Economic Review* 133, 1281–1302.
- Rabin, M., 1998. Economics and psychology. *Journal of Economic Perspectives* 36, 11–46.
- Rabin, M., 2000a. Inference by believers in the law of small numbers. Mimeo.
- Rabin, M., 2000b. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica* 68, 1281–1292.
- Rabin, M., 2002. A perspective on psychology and economics. *European Economic Review* 46, this issue.
- Rabin, M., Vayanos, D., 2000. The hot-hand fallacy and belief in the law of small numbers. Mimeo.
- Sartre, J.P., 1953. *The Existential Psychoanalysis* (H.E. Barnes, translator). Philosophical Library, New York.
- Schelling, T., 1978. Egonomics, or the art of self-management. *American Economic Review Papers & Proceedings* 68, 290–294.
- Smith, A., 1759. *The Theory of Moral Sentiments*. Liberty Classics, Indianapolis.
- Strotz, R., 1956. Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies* 23, 165–180.
- Thaler, R.H., 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1, 39–60.
- Tversky, A., Kahneman, D., 1971. Belief in the law of small numbers. *Psychology Bulletin* 76, 105–110.

- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 1124–1131.
- Weinberg, B., 1999. A model of overconfidence. Mimeo., Ohio State University, August.
- Williamson, O., 1964. *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm*. Prentice-Hall, Englewood Cliffs, NJ.