Introduction

ISABELLE BROCAS AND JUAN M. CARRILLO

The main reason for organizing a conference and editing a book that collects the views of both psychologists and behavioral economists is our intimate conviction that researchers in these two fields can and must work together. First, they can work together because both want to address the same kind of issues. Highlighting the connections between the different contributions is fairly easy, and part of this introduction is devoted to this task. The links are not exhaustive. Instead, it is just an indication of this potential for cooperation. Second, they must work together because, also as witnessed by the contributions in this book, similar questions are tackled from very different angles. There is indeed a strong complementarity between the two approaches that needs to be exploited, and cooperation seems the natural way to achieve this.

If the benefits of collaboration are so prominent, why do we observe some reluctance from both sides to join forces? In our opinion, the researchers are worried about the tools and methodologies employed in the other field. Naturally, there is no full agreement within each field, but still it is a mainstream view. Sometimes these concerns are well grounded, sometimes they are due to the development of research habits and the idea that a different perspective is necessarily less adequate. In any case, they constitute an important obstacle for debates and interactions. We would like to list a (incomplete) series of concerns that researchers in each field have for the practices of their counterparts. These have been gathered through informal discussions, mainly during the conference in which the papers collected in this book were presented. Obviously, the objections that can be made to any overgeneralization apply to our exercise. The reader should view this outline merely as an indication of some general attitudes and worries. Besides, the absence of input from a psychologist in writing this introduction increases the likelihood that we may have misinterpreted some arguments.[1]

One of the main concerns of some psychologists when looking at the behavioral Economics literature is what they consider an excessive degree of formalism. It is true that a certain amount of mathematical training is often necessary in order to understand theoretical contributions (and this may need some investment). However, the economists would argue that mathematical formulations are used exclusively as a communication language, whose advantages are clarity and precision. Second, the psychologists are also quite skeptical about the

[1] Needless to say that they also represent general views, which are not shared by every single researcher in the field.

'as if' approach of Economics where, provided that the models explain the facts, it does not matter whether the axioms are an accurate representation of the intrinsic features of individuals. The economists would claim that human decision-making is, for sure, governed by processes of a different nature than the underlying processes defined by economic models. However, this does not imply that the latter are not a good and dual representation of the former. Third, the economists are also criticized for their excessively simplistic view of the world and, in particular, of human nature. Obviously, the models are just caricatures of reality so, by definition, they only capture a tiny aspect of it. A blind belief on the prescriptions of a model is just as absurd as a straight rejection of its conclusions based on their simplicity. Fourth, there has been a tradition in Economics to believe that individual deviations from the standard rational behavior do not deserve a special attention, because the market forces alone will eventually correct these mistakes and/or displace these agents. Although this view is now challenged within the field, it has contributed to raise the suspicion of the psychologists about the economists' approach to human behavior.

How about the general concerns of economists about the practices in Psychology? Probably, the main criticism is the absence of a general framework in which all the contributions can be embodied. The economists like to think of papers as the pieces of a grand puzzle, which is in continuous construction, evolution, and even destruction. Furthermore, this general framework is supposed to provide a benchmark for comparison and a reference to be challenged by new theories and hypotheses. Naturally, a general approach to human behavior requires important simplifications and therefore goes against the desire of psychologists to provide an exhaustive treatment of every specific situation (see the third concern expressed above). For this reason, the psychology literature is more prone to adopt a case by case approach. Overall, there is an obvious tradeoff between accuracy and the generality of explanations. Another concern (which in fact is the opposite of the fourth issue raised in the previous paragraph) is the tendency in Psychology to extrapolate individual anomalies to the aggregate level. The economists argue that, even if deviants survive when the whole economy is taken into account, they are nevertheless likely to play a less important role. Naturally, the importance of this issue depends on whether the goal is to explain an anomaly or to quantify its consequences.

The absence of a unified vocabulary is another obstacle to inter-field debates, although this is probably somewhat easier to surpass. Notions like 'rationality', 'utility', 'preferences', 'emotion', 'cognition', 'esteem', etc. take different connotations for different people and this may result in severe misunderstandings. Ideally, in order to obtain a fruitful collaboration, people should share the way of formalizing problems. If this cannot be achieved, there should be at least a common language to define them.

Where do we go next? As Caplin and Leahy point out, the most productive alliances lie where strength meets strength. Probably, the main comparative advantage of the psychologists is their deep understanding of the behaviors,

feelings, and motivations of individuals, together with some well-developed experimental skills to test the competing hypothesis about human nature. The economists excel at developing normative frameworks, which can be used as a benchmark to conduct welfare analyses and obtain policy implications. By joining forces in this way, Economics and Psychology could confront the two major objections that each field has encountered separately. These are, respectively, an unrealistic degree of sophistication or rationality of individuals, and a lack of prescriptions to improve future behavior and increase welfare.

We will now present a synopsis of each contribution, placing a special emphasis on the links between the different papers, both in terms of the issues addressed and the methodologies employed. We hope that it will serve as a guide for readers interested in specific topics, but also that it will trigger-off their curiosity for other questions. However, first we would like to say a few words about the issues of rationality and irrationality. Probably, one of the most controversial debates concerning human behavior is to determine whether individuals maximize their expected utility or if they sometimes pursue self-defeating goals. In other words, how rational humans are. The fact that almost all the papers in this volume treat, in one way or another, this problem is a clear symptom of its importance. Nevertheless, the objective of this book is not to provide a straight answer to this question. Instead, each paper suggests some resolved or unresolved puzzles that should encourage further thinking.

I. THE CAUSES AND CONSEQUENCES OF 'IRRATIONAL' CONDUCTS

Since understanding the rationality or irrationality of choices is at the heart of the debates between the social psychologists and the behavioral economists, it seems natural to start this book with the contributions that address most directly this question. We might wonder whether it is important to know if a specific conduct can be termed as rational or not. Although each author has his own views about rationality, one can deduce from the papers that they are not so much concerned with the label itself. In fact, this issue is very much left to the reader's judgment, and we sympathize with this attitude. What really interests the authors are the reasons and repercussions of any kind of self-defeating behavior. The three approaches to this question collected in the book are introduced in the next paragraphs.

According to Baumeister (Chapter 1), three main reasons may push individuals to undertake self-defeating activities: deliberate self-harm, harmful decisions as an inevitable byproduct of some beneficial actions (and whose cost outweighs the benefits), and choices directed to obtain benefits but undertaken in a way that they produce the opposite effect. The author argues the existence of evidence only for the last two types of self-defeating behavior, and explores some theories to explain the circumstances under which people make such irrational choices. First, under emotional distress individuals are both less able to correctly evaluate the

consequences of their choices, and less likely to think carefully about the relative advantages of the different alternatives. This provides a rationale for the typical 'think twice before you act' maxim. Second, social considerations are also a cause of self-defeating actions. People with high self-esteem are concerned with the impressions made on others. When their ego is threatened, they are likely to make foolish choices. Similarly, feeling rejected partly inhibits the willingness to take care of oneself.[2] Last, Baumeister argues that people need some kind of energy in order to think correctly and that, as for any muscle, this resource depletes when it is used. Hence, forcing individuals to address a number of important issues may impair their subsequent ability for self-regulation.

Some theories about the cost of thinking have already emphasized the idea that humans cannot always evaluate all the alternatives before making their choices (see also the framing problem in Chapter 9). Baumeister takes a further step by arguing that the cost of thinking for each decision increases, as more and more choices have to be made. Moreover, if thinking refers to past events, this also provides support for the optimality of memory management, a theory that is carefully analyzed in Chapter 8. Pursuing this concept of self-regulation capacity as a muscle or form of energy, one may argue that the aptitude for self-regulation can be developed through its exercise. Therefore, if this conjecture is correct, self-regulation would then be costly in the short term but highly beneficial in the long term. We would also like to mention the striking ability of Baumeister (and, in fact, of all the authors in the experimental contributions of this book) to alter the moods and emotions of subjects for the purpose of their experiments, without raising suspicion.[3] For economic theorists like us, noting this capacity to create affective conditions is extremely reassuring.

Probably the most special contribution from the point of view of the economists and the behavioral psychologists is the paper by Berridge (Chapter 2), which looks at rationality from the perspective of a neuroscientist. The author starts with a specific definition of irrational behavior that builds on the distinction between liking and wanting: a choice is considered irrational if the individual chooses what he does not expect to like (or, equivalently, when the alternative selected is wanted disproportionately to its expectation of being liked).[4] Berridge argues that such a form of irrationality does exist. For example, a series of experiments on consumer behavior shows that the subliminal presentations of happy (resp. angry) faces, even though not consciously perceived, increase (resp. decrease) the rating and willingness to pay for a consumption commodity. The key issue is that the experimental subjects do not experience any change in their affective state. Hence, the decision utility or manifested choice of

[2] The key of Baumeister's argument is that choices under emotional shocks or ego threats are different from those under normal circumstances and yield a systematically lower utility.

[3] Chapter 13 deals in great detail with the issue of suspicion in experiments by the psychologists and the economists.

[4] Note the importance of the term 'expectation' since, in a world of uncertainty, it can be perfectly rational to (ex-post) regret and make (ex-ante) an optimal choice.

an individual can be altered by subliminal stimuli without affecting his predicted utility or expected liking for the commodity in the future (see Chapter 10 for a further analysis of the distinction between decision and predicted utility).

A natural issue for an affective neuroscientist is, then, to explore which brain mechanisms underlie aspects of decision and predicted utility. One well-known hypothesis is that the mesolimbic dopamine system mediates pleasure. However, this notion of pleasure refers indistinguishably to what the subject 'wants' and what it 'likes'. If, as Berridge's theory suggests, wanting a reward and liking a reward are two separate things, it may be that different brain systems activate these two psychological processes. This would in turn explain the mechanism behind the irrational conduct of wanting something more than it is expected to be liked. This hypothesis is tested in another set of experiments developed by the author. The main conclusion is that the dopamine activation affects decision utility (how much a reward is wanted) but not predicted utility (how much it is liked).

This theory has important policy implications. It is commonly argued that people consume addictive substances because they underestimate the probability of getting hooked. Once the individual is at this stage, a drug therapy that includes subliminal messages will not affect how drugs are perceived or liked, but it will decrease the incentives to consume them. This, in turn, will diminish the physical addiction component and therefore have long-lasting effects. Another interesting conjecture is that in stressful situations, predicted and decision utilities tend to diverge. Is it because stress also activates different brain systems? Note that this point is related to Baumeister's description of the set of situations in which individuals are likely to undertake self-defeating activities (Chapter 1).

Schooler et al. (Chapter 3) analyze the impact on actual happiness of both monitoring and pursuing happiness. In a first experiment, subjects are asked to listen to music for a period of time. The authors show that checking and reporting the moment-to-moment pleasure derived from the activity decreases its benefit. Similarly, attempting to maximize happiness leads to frustration that undermines the ability to feel the pursued pleasure. The overall conclusion is that monitoring and pursuing happiness are both self-defeating. A second natural experiment focuses on the ex-post assessment of happiness derived from a special event like New Year's Eve 2000. The conclusions are quite consistent with the previous experiment. They suggest that high prospects of a hedonic experience result in disenchantment. Moreover, the likelihood of feeling disappointed is proportional to the time and effort spent in the preparation of the event.

The paper is in line with previous research about the role of introspection in the ability to derive pleasure. As in this literature, introspection modifies the assessment of a hedonic experience. Besides, the self-defeating result is quite robust; it holds both for the pursuit and the monitoring of happiness, given a moment-to-moment and an ex-post evaluation of pleasure, and in a natural and a laboratory experimental context.

In our view, the results of the second experiment are far more striking than those of the first one. After all, thinking about one's feelings is costly (see

Chapters 1, 8, and 9 for ideas along this line). It may then be natural that the necessity to constantly provide a report diminishes the individual's capacity to experience pleasure. Instead, in the new Year's Eve case, the fact that only one report is asked implies that, according to the standard rational framework, the assessment should not be systematically modified. Indeed, there is a fundamental difference between reporting the pleasure of a past event (which measures remembered utility) and reporting moment-to-moment pleasure (which measures experienced utility).[5] Similarly, forcing individuals to pursue happiness (as in the first experiment) is also effort consuming, and therefore may decrease the total utility. By contrast, in the second experiment, subjects freely choose their level of investment in the pleasurable activity. Hence, those who decide to incur high costs should not be necessarily more disappointed than the others. This systematic bias in the expectations of hedonic experiences are closely related to the tendency of individuals to overestimate the likely duration of happiness following a positive event, a result obtained by Wilson, Gilbert and Centerbar, which is carefully explained in Chapter 11.

II. IMPERFECT SELF-KNOWLEDGE AND THE ROLE OF INFORMATION

The two obvious limitations of the traditional economic analyses of human behavior are the assumptions that individuals have an accurate knowledge of their own preferences and that their utility is not affected by the anticipation of future events. These are clearly unrealistic. For instance, the joy of future parenthood, the stress preceding a serious medical operation, or the tolerance to an addictive substance are not only hard to evaluate before actually experiencing them, they also influence the current well-being. But the lack of a precedent is not the only reason for an inaccurate self-knowledge. The temporal evolution of preferences and unpredictable changes in mood imply, that a repeated exposure to the same situation does not guarantee a convergence to complete self-knowledge and optimal behavior. Furthermore, as argued by Wilson et al. (Chapter 11), individuals are endowed with several self-regulatory mechanisms that systematically bias self-perception and therefore block the learning process. The papers presented in Part II study the three different aspects of the strategic value of acquiring and transmitting information, when either the current utility is affected by the anticipation of future events (Chapter 4) or self-preferences are imperfectly known (Chapters 5 and 6).

Caplin and Leahy (Chapter 4) argue that the anticipation of future events stimulates some emotions such as joy, excitement, fear, stress, apprehension, etc. These sensations have a direct impact on the current utility of individuals. Under these circumstances, obtaining news about the likelihood of future occurrences not only affects future welfare (through the standard effect of acting

[5] For the interested reader, Kahneman's contribution (Chapter 10) is fully devoted to this distinction.

under better information) but also current welfare, as uncertainty is a key factor that determines the intensity of the above mentioned emotions. When news is expected to stimulate negative emotions, avoiding information can be profitable.

The authors illustrate the effect of anticipatory feelings with the anxiety provoked by an upcoming surgical procedure. According to several studies reported in their paper, accurate information about the risks of an operation can, other things equal, either increase or decrease the anxiety of patients depending on the nature of the operation and the personality of the patient. Naturally, there are many other less dramatic situations in our everyday life where individuals have a strict preference for either early or late uncertainty resolution: the sex of a child before his/her birth, the weather forecast in the destination for the holidays, the physical aspect of a blind date, etc.

According to this theory, the beliefs of individuals about future events affect their current utility in a similar way than memories of past episodes. However, one important difference is the type of devices employed by individuals to increase their welfare. In the case of anticipatory feelings, Caplin and Leahy (Chapter 4) highlight the importance of information gathering and information avoidance. As we will see in Part III, in the case of bounded memory and imperfect recall, Benabou and Tirole (Chapter 8) and Gilboa and Gilboa-Schechtman (Chapter 7) develop theories centered around the strategic value of memory management and mental accounting, respectively.

Overall, in Caplin and Leahy, beliefs enter the utility function directly via the emotional predisposition. Given that information about the likelihood of future events affects not only future decision-making but also current payoffs, the anticipatory feelings raise the welfare issue of optimal provision of information. For example, should a doctor accurately inform his patient about the dangers of the upcoming operation? If the patient absorbs information at its face value, then the optimal strategy would probably consist of providing only stress-reducing news. However, patients are likely to anticipate the strategic provision of information by the doctor, that is to interpret the absence of evidence as stress-inducing news. Under these circumstances, the doctors might benefit from developing strict rules of behavior. All these welfare effects are important, especially if we take into account that moods have also a direct impact in the post-operation pace of recovery.[6]

Another natural situation where information can be detrimental, is the case of an individual with self-conflicting goals. Brocas and Carrillo (Chapter 5) study the optimal acquisition of information by an agent with imperfect self-knowledge, who discounts events in the near future relatively more heavily than events in the distant future. This type of temporal preferences—also known as hyperbolic discounting or present-biased preferences—has recently received a great deal of

[6] Note that the situations where uncertainty per se affects utility are pervasive. For example, many people declare that they do not enjoy watching a recorded sport game, even if they do not know the final score.

attention (see e.g., the references in Chapters 5 and 8 of this book). Probably, the main reason for the success of hyperbolic discounting is its greater predictive power of individual behavior than the traditional time-consistent exponential discounting. Obviously, such discounting is not suitable for every single situation; Trope and Liberman (Chapter 12) for example, develop a radically different theory of time-inconsistent choices.

Under the above mentioned conflict of preferences, Brocas and Carrillo show that an individual may optimally decide to avoid information, even if the news is freely available. As usual, being better informed allows him to improve his current decision-making. However, every piece of information is automatically shared with the future incarnations, who also take better decisions from their future perspective. Because of the intrapersonal conflict, these future optimal choices do not necessarily coincide with the first-best ones from the current viewpoint.[7]

The paper analyzes a wide range of situations in which the individual might be willing to remain strategically ignorant. It provides testable predictions concerning the aggregate effects of rational ignorance on observed behavior. If the activity involves current benefits and negative internalities (i.e. costs on future welfare), the individual will hold pessimistic prospects about the expected payoff of undertaking the activity. By contrast, if the activity yields current costs and positive internalities (i.e. benefits on future welfare) he will remain deliberately optimistic. An illustration of this idea is the case of smoking (a pleasurable activity with negative internalities). The fear of overconsuming cigarettes due to hyperbolic discounting, may induce an agent with negative current beliefs about the effects of tobacco on his health not to collect extra information and quit. Conversely, an individual facing difficult research projects (that require substantial effort but provide positive internalities) will hold positive beliefs on his capacity to succeed in order to avoid inefficient procrastination.

One interesting alley for future research would be to combine the preferences over beliefs motive for information avoidance (Chapter 4) with the intrapersonal conflict motive (Chapter 5). In particular, one may wonder in which direction the inclusion of anticipatory feelings on the incentives of an agent with present-biased preferences to acquire information is likely to affect.

The paper by Bodner and Prelec (Chapter 6) also deals with the issue of imperfect self-knowledge. The authors study optimal decision-making by an individual with different levels of conscience. According to their theory, when an individual undertakes an action, he signals his preferences not only to the outside world but also to himself. For example, a dieter unable to resist the temptation of an extra piece of cheese cake reveals to himself a lack of willpower, an individual who starts an extramarital relation draws inferences about the true love for his partner, a night spent in a casino signals to a gambler his propensity to this vice (this last example is studied in detail in the paper), etc.

[7] Benabou and Tirole (Chapter 8) include imperfect memory in this same framework. As a result, every piece of information is not necessarily shared.

Naturally, a necessary condition for actions to be a diagnosis of preferences is that the information possessed by the agent when making decisions must be greater than when interpreting the effects of such decisions. In Bodner and Prelec's words: 'the gut knows [...], but the mind does not'. Hence, in this context, the reason for the agent's willingness to manipulate information is not the existence of an intrapersonal conflict of preferences (as in Brocas and Carrillo) but rather the possibility of drawing inferences about oneself from the selected actions.

One of the most important issues of this self-signaling approach is the degree of sophistication (or rationality) of the individual. Does the decision-maker take the action anticipating the diagnostic effects of his choices, that is the impact of the decision on his future self-perception and utility? The first possibility is to assume that the decision-maker only maximizes his utility. Choices are then interpreted at their 'face value' and inferences are just a byproduct. One example is the dieter who refrains from having dessert depending on the strength of the temptation, and happily concludes that he possesses a great willpower whenever he succeeds in controlling his desires. A second possibility is to assume that the individual integrates the effects of his actions in the revision of beliefs. Under this approach, the same dieter knows that he might have resisted the temptation, precisely to convince himself that he has a great willpower. Although closer to individual rationality, this alternative is nevertheless paradoxical: how can someone draw a conclusion from an action whose goal is precisely to achieve this inference?

We would like to conclude by mentioning some considerations that could be of interest for further investigation in a self-signaling framework. First, the individual may decide to seek or avoid the type of situations in which the unknown trait is likely to be revealed, depending on the anticipation of his reaction. For example, the willpower of a dieter will be especially threatened at family dinners (this point is addressed by Benabou and Tirole (Chapter 8) in a context of the hyperbolic discounting of future payoffs). Second, the importance of the diagnosis effect of choices is endogenous. The value of learning about self-preferences is given by the likelihood of facing a similar situation in the future, but this probability itself affects the current behavior. Last, two (Bayesian) individuals with the same expected beliefs may exhibit radically different attitudes if these beliefs are drawn from different distributions: an uncompromising individual (who interprets any relapse as evidence of a complete lack of willpower) will adopt a more rigid attitude than a more self-indulgent person. Including some of these possibilities would increase the number of insights that can be gained with this novel approach to human reasoning.

III. IMPERFECT MEMORY AND THE LIMITED CAPACITY TO PROCESS INFORMATION

We have emphasized in Part II the role played by the anticipation of future feelings and actions on current decision-making. The memories of past events have a mirror image effect on behavior. Surprisingly, except for some notable

exceptions, economic analyses have largely neglected the natural tendency to recall imperfectly past information and decisions. One standard reason for not incorporating this human weakness, is to argue that individuals will keep track of past news or behaviors if these are sufficiently important. For most everyday situations, it seems clear that recording activities is not an easy task, as witnessed, for example, by the popularity of complicated electronic memory devices such as the Palm Pilots. The three contributions introduced here study the different aspects of an imperfect ability to analyze information. A distinction is made between imperfect recall (Chapters 7 and 8) and imperfect information processing capacity (Chapter 9).

The starting point of Gilboa and Gilboa-Schechtman (Chapter 7) is the observation that, in some circumstances, individuals take current decisions as if they create a precedent for future conduct. Naturally, under free will, the application of this Kantian's categorical imperative to oneself constitutes a cognitive mistake. The paper argues that an individual with imperfect recall may not be able to remember all the contingencies that led to his past decisions. If this is the case, then a simple rule of behavior expected to be blindly imitated by every future incarnation may in fact be the optimal strategy. The authors illustrate this idea with several examples of such rules: 'buying an expensive sweater only for my birthday', 'taking two glasses of wine only in family dinners', etc. These behaviors allow some extravagant purchases and some indulgence of vices, but they are not dangerous for the financial and health states, respectively.

One should notice that the type of situations analyzed above belong to a broad class of phenomena termed in the psychology literature as 'mental accounting'. According to this theory, any decision can entail a psychological side effect that cannot be ignored by the individual, but which may or may not operate depending on the circumstances (following the previous example, a person might not feel guilty when buying an expensive sweater, if and only if it occurs under an exceptional occasion). Usually, the mental accounting literature takes for granted this psychological reasoning, and limits the attention to analyze its costs and benefits for the individual as well as its effects on decision-making. The strength of the paper by Gilboa and Gilboa-Schechtman is to show that this heuristic 'overgeneralization' conduct may, in fact, be the result of an optimization process under the constraint of imperfect recall.

Although it is a theory based on imperfect memory, one can notice the intimate relation of Chapter 7 with the papers by Bodner and Prelec (Chapter 6) and Brocas and Carrillo (Chapter 5). As in the first one, the individual uses his own actions to convey information to himself (useful because of bounded memory rather than because of an incomplete information with oneself). As in the second one, the individual achieves some self-discipline (through the self-imposed rule rather than through strategic ignorance), which is beneficial for his long term payoffs. Last, the paper raises a number of puzzling questions that could be the object of further study. Of particular interest would be to investigate the mental processes that determine in which directions the overgeneralization is formed.

Benabou and Tirole (Chapter 8) explore further aspects of imperfect recall. Their theory adds the limitation of imperfect recall to the agent's ability to retrieve information (common to the other contributions in Part III), to three other aspects that are non-standard in traditional economic models. Two of them, imperfect self-knowledge and the hyperbolic discounting of payoffs, have already been discussed. The third one, introduced to account for selective memories, is the ability to influence the probability of remembering past events. However, it is noted that individuals cannot fool themselves, that is, they cannot decide which information is retained and which one is forgotten. The paper reviews a surprisingly rich set of predictions obtained with a combination of these elements.

One particularly interesting result is that individuals may exhibit either over- or under-regulation of their behavior. While deficient self-restraint has been often emphasized as the result of a hyperbolic discounting of future payoffs, these type of preferences have long been criticized for their inability to explain compulsive, excessive self-restraint behavior (like greediness or workaholism). Benabou and Tirole show that such conducts are just 'two sides of the same coin' and that, depending on the initial beliefs about one's capacity to resist temptations, individuals may exhibit one attitude or the other. When memory repression is costly but recall is costless, the authors show that the incentives for memory manipulation are higher the degree of time-inconsistency and the lower the cost of repression. Last, Benabou and Tirole also analyze (in a different context and without issues of imperfect memory) some problems related to social interactions. They consider relations like parents/children and teachers/students. The former have some private information that is important for the decision-making of the latter and may decide whether to offer explicit incentive schemes, in which good performance is rewarded. The paper argues that, in some circumstances, such schemes may be weak positive reinforcers. Furthermore, such rewards become negative reinforcers once they are withdrawn.

In our view, one of the strengths of the memory management theory developed in the paper is that it endogenizes the individual's level of recall within the traditional Bayesian framework.[8] The authors highlight the fact that the hyperbolic discounting agents are more willing to repress memories than their exponential discounting peers. However, if recall is also costly, the marginal benefits of remembering past episodes is sometimes smaller for the former than for the latter. Overall, our intuition is that predictions about which type of agent is going to exhibit the highest degree of recall will not always go in the same direction. Another important point is that, as the paper argues, affecting the probability of remembering some episodes might not have the expected impact on behavior and welfare. Trying to forget may well be an attempt to escape what cannot be escaped. Psychoanalysts and psychotherapists base their therapies on the fact that

[8] Obviously, one should not take this approach too literally: individuals do not decide on how much to recall. However, the probability of not forgetting can be increased (memory exercise, written notes, etc.) or decreased (attention management, drinking, etc.).

individuals never forget important (positive or negative) episodes of life; memories can be repressed in the unconscious, but they still affect behavior. It would be interesting to find some support in the neuroscience literature, for this self-regulatory ability of individuals to affect their capacity to recall events.

More generally, Chapters 7 and 8 raise important questions about how the mind (and specially the memory) works. For instance, our ability to memorize episodes is likely to depend on the episodes themselves. In particular, there is some evidence that each episode of our life is stored in a particular area of our brain where other similar events are also stored, and all are consciously remembered in the same proportion. Is there one kind of memory or many? Where is memory stored? How is memory organized? Are there different kinds of storage processes that serve the maintenance of different types of information in memory? These questions are still open.

Gabaix and Laibson (Chapter 9) offer a completely different approach to the issue of an imperfect capacity to process information. They argue that, in complex problems, even a superpowerful computer cannot determine the payoff of all the different alternatives. As a result, humans employ some heuristic simplifications of the problems, in order to come as close as possible to the optimal solution. Their goal is, then, to understand which are the main tricks or rules of thumb that lie behind human inferences.

To explore this simplification process, the authors assume that attention and cognition are scarce resources (a point also made in Chapter 1) that have to be allocated as efficiently as possible. Then, they build a model based on the premise that the option of continuing the analysis declines when cognitive analysis yields little new insights, and when one particular choice gains a large edge over the other alternatives.[9] Once these theoretical option value postulates are established, it is possible to determine the optimal choice given this specific simplifying rule to reduce the complexity of the problem. The authors test their model by comparing the realized choices of experimental subjects in complex settings with the predictions of (i) their 'directed cognition' model, (ii) the rational choice model (in which the exact payoff of all the alternatives can be computed), and (iii) some other variants of the rational choice model (in which, for example, all the information in the distant future is disregarded). Interestingly, Gabaix and Laibson's model outperforms the rational choice model and (at least) breaks even with its variants in predicting the behavior of individuals.

This paper can be viewed as a first step towards a better understanding of the heuristics employed by individuals to think both quickly and effectively. Naturally, the next question is to determine how robust is its predictive power, when we introduce some modifications in the problem under scrutiny. For example, in the experiment considered in the paper, subjects choose between several decision trees each of them with different intermediate payoffs, but of the same order of

[9] Stated differently, thinking is more valuable when there is a lot of information to be learned, and when there is much more information to be learned in one alternative than in the others.

magnitude. One may wonder how the predictive power of the model would be modified if one alternative were substantially better than the others, but the way to reach it remained uncertain. The authors also argue that a major benefit of understanding which are the most common simplification tricks is that, this would help to determine the circumstances where humans are likely to commit systematic judgment errors. Naturally, the prescriptions to avoid such mistakes would then be tremendously helpful. Last, in many cases we are not even able to determine ex-ante, which are the relevant alternatives, and instead these are discovered when some intermediate decisions are already made. It would be interesting to determine whether a similar type of model would be a good predictor of the agents' behavior in this alternative setting.

IV. TIME AND UTILITY

As we saw in Part II, the temporal dimension can affect the assessment of utility if the individual has anticipatory feelings (Chapter 4) or an intrapersonal conflict of preferences over time (Chapter 5). However, time and utility are connected in many other subtle ways, through the mental representation of future events and memories of past episodes. The papers presented here offer three new theories about these interactions.

Kahneman (Chapter 10) proposes a new way of evaluating the utility of individuals over a period of time that is the starting point for the conception of an objective measure of happiness. The author starts from Bentham's notion of experienced utility (the idea that utility refers to past experiences that in turn determine the optimal future decisions to be made) and extends the concept to an intertemporal setting. He claims that a 'memory-based approach' that depends on a retrospective assessment of the whole experience is often a bad way to assess the individual's overall experienced utility. For example, in the case of a painful experience with a varying intensity, the total remembered utility depends mostly on the highest level of pain and the intensity of pain suffered at the end (the peak/end rule). In particular, duration has only a mild effect in the evaluation of utility. Kahneman argues that, under some conditions, a 'moment-based approach' in which total utility results from summing all the instantaneous moment-utilities is the best measure of the overall experienced utility. Besides, the moment-based approach is also suitable for constructing a measure of objective happiness. This measure has the important advantage over standard subjective well-being criteria that, it does not require the individuals to report their memories of past experiences.

The paper raises a number of important questions and policy implications, some of which are partly addressed in the discussion section. One issue of interest is to understand the causes and consequences of distorted memories about past utilities. It may be that adverse past experiences have a negative impact on current and future pleasure, so that the distortion is just a self-regulatory mechanism that helps people to overcome such experiences. A consequence is that it induces

individuals to repeatedly incur the same mistakes in their decision-making (naturally, from a moment-based point of view !).[10] A second point of interest is to compare the essence of moment-based and memory-based utilities. As Kahneman points out, the present experiences are the only real ones. This suggests that the moment-based approach captures the aggregation of real experiences. On the other hand, what determines our current feelings and moods are the memories of the past and the anticipations of the future. These are best captured with the memory-based approach.

The theory certainly has many possible applications. Consider for instance the following provocative example. The main goal of imprisonment is not to make the convicts suffer but rather to deter them (ex-ante) from committing a crime and (ex-post) to relapse.[11] Applying the previous terminology, the optimal policy is then to maximize the (negative) remembered utility and minimize the (negative) total utility. These two goals can jointly be achieved by inflicting short punishments with an intensive painful peak. Furthermore, a measure like probation, which is basically a soft and prolonged punishment, should be entirely avoided.

We would like to conclude the review of this paper by noting that using the difference between remembered and total utility to design policies, raises serious ethical concerns. In particular, as the author shows, it is fairly easy to manipulate people's decisions by making experiences intensively painful or pleasurable at some points in time. It is also unclear how to propose welfare prescriptions. Indeed, even if they are manipulated the people feel happy—in which case manipulation could be viewed as improving welfare—,nevertheless these individuals have their will restricted. Furthermore, it is not clear which utility criterion (total or remembered utility) is the most adequate to measure welfare.

The experienced utility may not only reflect the evaluation of past events but also the beliefs about the evaluation of future events, in which case it is termed as a predicted utility. The objective of the work by Wilson et al. (Chapter 11) is very much related to this notion of predicted utility. The authors study the ability of individuals to forecast their future affective state. The authors show that individuals are quite accurate in predicting the valence and the intensity of their emotional reactions. However, they suffer from a durability bias, that is a systematic overestimation of the duration of their future feelings after an affective shock. This bias occurs independently of whether the emotion is positive (winning a large sum of money in the lottery) or negative (death of a loved person).

Wilson and his colleagues show that the main reason for such an emotional evanescence is a human innate tendency to 'reduce [ex-post] uncertainty about the world by finding meaning in it'. According to this theory, the notion of

[10] Note also that, in order to address this matter, one should first determine whether such distortions depend on the valence of the experience.

[11] Some people argue that it is also a way of putting them aside in society. We will not discuss this point.

durability bias is then intimately related to that of hindsight bias, which is the tendency to find ex-post predictable, the realization of an ex-ante uncertain event. In other words, the rationalization of an occurrence decreases its emotional content, because being more predictable also means being less extraordinary.

The paper takes a further step by arguing that emotional evanescence can also be functional. As in the case of food, there is an optimal degree of happiness. In order to maintain the emotional state within some bounds, a self-regulatory mechanism is automatically triggered-off whenever it reaches an excessively low or excessively high level. Last, individuals may increase the permanence of pleasure by inhibiting the process of making novel events that look ordinary. As shown in some experiments reported in the paper, the emotional state of individuals is most durable when they are most uncertain about the reasons for the occurrence of a positive event.

To sum up, this paper explores in detail some reasons for the durability of happiness. It is therefore complementary to the contributions on the pursuit (Chapter 3) and the evaluation (Chapter 10) of happiness. It also raises some questions about the accuracy in the evaluations of future emotional states. One might argue that some individuals deliberately misreport their affective forecasts. This can occur either because of social norms (people would probably doubt my feelings as a parent if I state that life would go on even if my child dies) or as a self-justification of current choices (working excessively to get a particular job is justified only if I think that it is going to change my life afterwards). Last, it would be interesting to explore in more detail the costs of the durability bias. The authors insist on it in the case of negative events. However, even for a positive event, overestimating the duration of happiness can also be extremely harmful. For example, an individual may undertake an ex-ante costly investment decision with the anticipation of an immense expected benefit. Such a decision can be foolish if the pleasure ends up being mild.

The paper by Trope and Liberman (Chapter 12) adds a third perspective to the evaluation of events at different points in time, and their effect on utility. The starting point of their 'temporal construal theory' is the idea that the representation of a given activity depends on its temporal distance. In particular, when the distance from a future event is important, its representation is more likely to contain essential features (high-level construals) and less likely to contain superficial features (low-level construals) than when the distance is small. Hence, when the high-level construal of an activity is positive (resp. negative) and its low-level construal is negative (resp. positive), then it is more likely to be desirable than another activity with a neutral high and low construal as the distance for the realization of the activities increases (resp. decreases). One should notice that, the most fundamental difference in the traditional theories of time-inconsistent choices (such as the hyperbolic discounting theory presented in Chapters 5 and 8 or the valence theory) is that the temporal construal theory builds on changes over time in the *mental* representation of events, rather than on the mechanical changes in the relative discounting of consecutive periods.

The authors provide several experiments that support their theory. In one of them, some subjects were asked to indicate their interest in performing an interesting main task (i.e. with a high and positive construal) together with a boring filler one (i.e. with a low and negative construal). The other subjects had to indicate their interest in performing a boring main task with an interesting filler one. Not surprisingly, the individuals showed a greater interest for the first combination of tasks than for the second one. More important, the difference in the attractiveness of the two combinations was greater when the subjects were supposed to perform the tasks in the distant future than when they were supposed to perform them in the near future.

Overall, the main prediction of the temporal construal theory is that the distant future choices are mainly influenced by the central, superordinate, abstract (high construal) characteristics, whereas the near future choices are mainly regulated by the peripheral, subordinate, concrete (low construal) features. However, one should realize that the same activity can be *perceived* by the individual as having a high or a low level of construal depending on how it is presented to him, that is it can be manipulated to some extent. This has important implications. For example, working in an organization often implies performing a job that has both an easy, simple paperwork component and a difficult, innovative thinking component. Trope and Liberman's theory predicts that, if the task has to be completed in the near future the employer should motivate the individual to work hard on it by emphasizing ease. In contrast, if it has to be completed in the distant future, then motivation is best achieved by emphasizing its challenging aspect.

V. EXPERIMENTAL PRACTICES IN PSYCHOLOGY, ECONOMICS, AND FINANCE

Performing a controlled experiment is probably one of the most natural and direct ways to improve the understanding of human conduct. The psychologists have known this for a long time. Nowadays, the research papers published in leading behavior Psychology journals, where theories are not supported by some empirical or experimental evidence hardly exist. Surprisingly, this practice is much less common in Economics. Certainly, experimental Economics is a well-developed field, whose relative importance is constantly rising. However, there still exists a significant fraction of researchers (which includes the authors of this introduction) who devote their effort to present new theories (based on introspection, casual observation, previous empirical or experimental research, etc.), analyze their implications, and provide some policy prescriptions but do not offer any new test of their arguments. As a well-known behavioral psychologist not familiar with the practices in Economics put it: 'Many economists tell a nice story, provide an elegant model and draw convincing conclusions . . . but where is the data?' Some people may argue that a basic comparative advantage argument pushes towards specialization in either theory or experiments. However, it is pretty obvious that sometimes the only way to fully understand a problem is by

addressing both the issues simultaneously. The papers collected in Part V offer three different perspectives on experiments in Psychology, Economics, and Finance.

Hertwig and Ortmann (Chapter 13) compare the experimental practices in behavioral Economics and behavioral Psychology. They argue the existence of four main differences in the design of experiments. First, the economists provide a precise set of actions among which the subjects of the experiment have to choose; the psychologists often do not. Second, the economists systematically repeat experimental trials; the psychologists do not. Third, the economist almost invariably provides monetary incentives based on performance; the psychologists do not. Last, economists rarely deceive the participants in their experiments; the psychologists often do.

Their analysis focuses on the last two issues. It shows that, performance-based rewards are often important for inducing participants to incur the cognitive effort necessary to avoid errors in judgment. In other words, under monetary incentives, the behavior of individuals is closer to the predictions of normative models and the variability of data is reduced considerably. Concerning deception, the authors argue that anticipating the possibility of being deceived raises the suspicion and the second-guessing of the participants. This, in turn, distorts their motivation and ultimate behavior.

Naturally, these reasons are also well-known by psychologists. The main counterargument exposed for the suitability of avoiding monetary payments is that, the goal of many experiments is, precisely, to test the motivations of individuals to undertake certain actions. Strong monetary incentives crowd out other intrinsic motivations, and therefore bias the results. As for deception, there are two main reasons in favor of its use. First, when the study concerns issues like prejudice or social conduct, an open statement of the purpose of the study will induce subjects to act strategically and conform to what they believe is the socially acceptable conduct. Second, it is sometimes necessary to artificially create some conditions, which are otherwise unlikely to occur, in order to perform a controlled experiment. An extreme example is to produce an emergency situation in order to analyze decision-making under stress.

Overall, the authors argue that it is obviously impossible to include monetary payments and to avoid deception when non-monetary rewards and deception are precisely the object of the study. However, they also claim that experimenters should avoid these two practices whenever the goal of the study points towards a completely different direction. In other words, they advocate an empirical approach to the evaluation and selection of variables in experimental designs. In our view, there are a couple of points that deserve a more in-depth treatment. One is that 'monetary' and 'performance-based' payments are often different issues. First, because flat participation fees with no (or small) bonuses if results are good can be considered as monetary incentives, but they are not (or weakly) correlated with performance. And second, because it is possible to offer rewards that are contingent on performance without necessarily implying money transfers (high

grades, granting the participation in a more important study, etc.). The other is that, it is sometimes difficult to determine whether suspicion affects behavior or not: when the authors report a study in which suspicious participants have a less conformist attitude than their non-suspicious peers, it may well reflect the existence of a correlation between the two characteristics rather than a causality.

Hilton (Chapter 14) reviews the psychological traits, biases and judgment errors that are likely to play an important role in financial contexts. The paper starts with a description of individual actions and judgments that can be labeled as 'irrational' according to the traditional definition of the homo-economicus, as well as some experiments that support the existence of such a departures. The behaviors include illusion of control, overconfidence, a false interpretation of evidence, loss aversion, framing effects, and the underuse of information. The author then analyzes some social conducts that are also hard to reconcile with the traditional utility maximization paradigm: irrational herd behavior, over- and under-reaction to news, communication and inefficient aggregation of information, and the effects of accountability on decision-making. Last, the paper provides some potential applications of some well-known findings in psychology to the marketing of financial products and the management of human resources. In particular, it discusses how the design and presentation of financial products affects the decisions of non-professional investors, and how an efficient training of traders directed to correct their biases in judgments can increase their performance.

Hilton's contribution is a clear example of the potential benefits of a tight collaboration in the fields of Psychology and Finance. In recent years, there has been a growing tendency in the field of behavioral finance to rediscover the propensity of individuals to make systematic errors in judgment and decision-making. A more efficient research strategy would be to rely on the rich tradition that psychologists have in analyzing these patterns of behavior. Using the already existing conclusions, it is then possible to draw the inferences that are most relevant for the field and to provide policy prescriptions that improve the efficiency of institutions. However, one of the biggest dangers of adopting this strategy is the possibility of losing the rigorous and critical approach that economists have always displayed when analyzing human behavior. In other words, before invoking a non-standard argument to explain a given behavior, it might be useful to clearly understand whether and why the traditional theories do not provide satisfactory explanations to the issue under study.

One interesting alley for future research, not discussed in the paper, is to study the circumstances under which an irrational individual behavior is likely to survive in the long run. The social value of individual irrationality has been recognized for a long time. An example borrowed from the finance literature: an individual who completely disregards public evidence in favor of his own criterion reveals by the means of his actions his private information to the entire population. The private value of irrationality has also received some support. For instance, there is always one 'crazy' individual in every group or tribe. This

person usually gets the best deals in simple bargaining situations, independent of his/her true level of insanity. However, it would be interesting to provide a systematic analysis of the type of situations in which irrationality is likely to improve outcomes either for the individual himself or for the society as a whole.

The relation between irrational behavior at the individual and at the aggregate level is at the heart of the last contribution in Part V, namely the one by Fehr and Tyran (Chapter 15). As the authors argue, one of the main differences between the researchers in Psychology and in Economics or Finance is the following. On one side, the psychologists analyze individual anomalies and then extrapolate for the aggregate behavior. On the opposite side, economists overestimate the likelihood that competitive market forces correct individual anomalies.

The paper analyzes the problem of individual and aggregate irrational behavior in the case of 'money illusion'. This theory roughly says that illusion-free individuals should react to changes in real prices but not to changes in nominal prices (in other words, I should feel equally happy if both the inflation and my wage increase by 4 percent than if both remain constant). The authors propose a market experiment in which individuals post prices and obtain benefits, depending on the prices set by other individuals. The game is characterized by strategic complementarities (the optimal reaction to a price increase by the opponents is to increase own prices) and a nominal shock, that is a change in the quantity of money, which does not affect profits if all the individuals change their nominal prices in the same percentage. The paper shows that, not surprisingly, individuals exhibit more inertia in their price adjustment to shocks when prices are posted in nominal rather than real terms. More importantly, the adjustments are also slower when individuals play against other individuals than when they play against computers. The overall conclusion is that the severe nominal inertia is not so much due to individual irrationality but rather to the uncertainty of individuals on whether they are facing other rational individuals.

Although the paper is concerned with a very specific issue (of interest mainly for economists), its implications are much more general. In fact, taking a close look at the general experimental setting, one can realize that the results are indeed special instances of two widely recognized theories in psychology: the importance of framing effects (i.e. how different presentations of the same situation lead to different individual responses) and the role of beliefs about others' beliefs (just like in the famous 'beauty contest' example). It would be interesting to explore the interactions between these two phenomena in other contexts.

SOME CONCLUDING REMARKS

The main goal of this introductory chapter has been to provide some flavor of the fifteen contributions collected in the book. We have tried to emphasize the similarities in the questions addressed by the different authors and, at the same

time, the diversity in their approaches. In particular, as we have discussed, there are several marked differences between the methodologies employed by the economists and the psychologists. In our view, there is a tremendous potential for cooperation between researchers in the two fields due, precisely, to this combination of common interests and different perspectives.

We have grouped the papers according to some general themes such as irrationality, self-knowledge, memory, etc. However, there are so many similarities in the issues addressed by the authors, that it would have been equally possible to propose totally different topics, like time-discounting theories, happiness, emotions, etc. We also hope that the questions raised in the brief review of each chapter and the suggestions for further investigation will stimulate the intellectual curiosity of the readers.