Mental Accounting and the Absentminded Driver*

By
Itzhak Gilboa
Eitan Berglas School of Economics
and Recanati School of Business
Tel-Aviv University
and
Eva Gilboa-Schechtman
Department of Psychology
Bar-Ilan University

Abstract

There is a broad range of phenomena that are categorized as "mental accounting". Some of these phenomena may be viewed as situations in which a person evaluates decisions based on the implicit assumption that the decision she makes at a given moment will be repeated many times in the future. For example, people find it easier to spend windfall profits than they do regular income. One explanation might be that "always spending a lot of money" is a bad strategy, whereas "always spending windfall profits" is not.

The mental accounting phenomena that belong to this category can be described as involving modes of thinking in which a single agent (or "self") of the person follows Kant's categorical imperative with respect to her future agents ("selves"). Namely, despite the fact that the environment does not provide any reason to believe that a causal relationship exists between the decision of the agent in question and those of future agents, the former does evaluate decisions as if the latter were guaranteed to follow her example.

How can one explain such a cognitive mistake? We argue here that one such explanation is provided by imperfect recall. As the term is used in game theory, it suggests that an agent of a player might not recall what a past agent of the same player has done, or even whether it existed. In these situations, one may think of the player as giving identical instructions to all future agents: since these agents can't even identify themselves, they cannot be given separate instructions. Under this constraint, it may be optimal for the player to give her future agents instructions that exhibit mental accounting effects.

* We wish to thank Michele Piccione and Drazen Prelec for comments and conversations. We are very grateful to Juan Carillo for many specific comments. Israel Science Foundation grant no. 10110751 is gratefully acknowledged.

1. Mental Accounting and Overgeneralizations

The term "mental accounting" refers to a variety of phenomena in which money is non-fungible. Since the pioneering works of Kahneman, Tversky, and Thaler (Kahneman and Tversky (1979, 1983), Tversky and Kahneman (1981) and Thaler (1980, 1985)) it has become abundantly clear that people do not treat all dollars alike. Rather, people manage different accounts in their minds, in which money has origins and goals that cannot be ignored when this money is spent.

There are many types of mental accounts and various rationales for drawing distinctions between seemingly identical resources. This chapter proposes that one type of mental accounting has to do with overgeneralization, and that it can be viewed as an optimal strategy under a constraint of failing memory. The following example illustrates.

Example 1 (Thaler (1985)): Mr. S admires a $125 cashmere sweater at the department store. He declines to buy it, feeling that it is too extravagant. Later that month he receives the same sweater from his wife for a birthday present. He is very happy. Mr. and Mrs. S have only joint bank accounts.

What goes on in Mr. S's mind when he declines to buy the sweater and when he later on receives it as a gift? It might be the case that, when Mr. S receives the gift from his wife, he is relieved of the moral responsibility that would normally be attached to consumption decisions. That is, if *he* makes the decision to buy the sweater, he should pay the mental and social cost of being a spendthrift. Having made such a decision may not conform to his self-image as a frugal person; it may also serve against him in his attempt to control his teenage daughter spending, and so forth.

An alternative explanation is that Mr. S feels that it is acceptable to buy the cashmere sweater for his birthday, but unacceptable to buy it with no other reason apart from the fact that the sweater was there. One may consider two other scenarios that might distinguish between the two explanations: first, Mrs. S may buy the sweater as a gift for Mr. S with no special occasion or special reason in mind. Second, Mr. S might consider buying the sweater himself for his own birthday.

Mental accounting theory suggests that control over and responsibility for choice are not the only reasons that money may differ from money. Thus, Mr. S may still feel uncomfortable with the sweater if it was bought by his wife without the birthday excuse, and may also feel that it is appropriate for him to indulge himself on the occasion of his birthday and buy the sweater himself. In both situations, Mr. S may say to himself, "Well, I couldn't possibly walk into a store and grab whatever I like whenever I like it, but I can afford to spend a bit more once a year." That is, according to this particular explanation, Mr. S's assessment

of his choice does not depend only on the outcome of the present decision problem, but also on the hypothetical outcome of following the same mode of decision making in many decision problems to come. It is as if Mr. S overgeneralizes his decision. Rather than asking himself "Is it OK for me to buy this sweater", he asks, "Is it OK for me *always* to buy things like that?" Obviously, "things like that" is a very vague term. Indeed, propositions can normally be generalized in many different ways. The proposition "Today Mr. S is buying a $125 cashmere sweater" may be generalized to "Mr. S always buys $125 cashmere sweaters", to "Every Monday Mr. S buys expensive clothes", or to "Mr. S always buys whatever he fancies". Which generalization should Mr. S use in assessing the current decision? It is only natural that the details of the purchase decision will suggest what is the most natural generalization. Specifically, if Mr. S has no special reason for buying the sweater on this day, a rather sweeping generalization suggests itself. By contrast, if the sweater is a birthday gift, it is quite natural to think about a future in which "Every year Mr. S spends $125 on a birthday gift." Clearly, the second generalization is financially much less problematic than the first. Correspondingly, the prospect of spending $125 once a year on a birthday present is reasonable enough to allow Mr. S to enjoy the sweater. But the prospect of spending $125 every day is rather scary. So scary, that Mr. S would rather decline to purchase the sweater.

Let us assume that people do indeed tend to overgeneralize and to evaluate choices as if these choices are bound to be repeated in similar situations in the future. Based on this premise, one may explain several examples of mental accounting. Consider the following.

Example 2 (Thaler (1980)): A family pays $40 for tickets to a basketball game to be played 60 miles from their home. On the day of the game there is a snowstorm. They decide to go anyway, but note in passing that had the tickets been given to them, they would have stayed home.

Why would this family insist on using tickets they have bought but allow themselves to forego the game in case the tickets were given as a gift? Assuming that they judge decisions based on the assumption of recurrence, the two situations turn out to be rather different. In one situation the generalization is "We always/often buy tickets we do not use" whereas in the other – "We always/often do not get to use tickets we have been given". The first generalization is financially ominous. There is no bound on the sum of money one might lose if one were to buy tickets without using them. By contrast, one normally would not go bankrupt simply by failing to make the best use of gifts.

Example 3: (Thaler (1985)): Mr. and Mrs. L and Mr. and Mrs. H went on a fishing trip in the northwest and caught some salmon. They packed the fish and sent it home on an airplane, but the fish were lost in transit. They received $300 from the airline. The couples take the

money, go out to dinner and spend $225. They had never spent that much at a restaurant before.

Again, one possible explanation of this behavior is that the couples judge the choices available to them under the assumption that they will be repeated in the future. Spending $225 at a restaurant without the fish story suggests generalizations of the type "We spend large amounts of money out of our income when we feel like it". By contrast, spending the same amount in the story above generalizes more naturally to "We spend windfall gains on luxurious items". Whereas the former generalization hints at financial ruin, the latter bodes no ill. Hence the $225 meal can be enjoyed in the story above, whereas the same meal without the fish story would taint pleasure with concern for the future.

Observe that in the example above the couples spend the money on a meal, rather on some other luxurious consumption. Namely, they do not only have a mental account for "easy come, easy go" money, but they seem to have an account dealing more specifically with food or with their lost salmon. The linkage between the source of the money (compensation for lost salmon) and its use (a good meal) is reflected in the generalizations that are likely to be generated. If, for instance, the couples were to buy an expensive coat with the money they received from the airline, the effect of the mental account might have been diminished. This may be due to the fact that spending the airline money on a coat might be generalized to "Spending money (that came from a generic source)" more easily than in the present example.

Undoubtedly, there are other explanations for these examples of mental accounting. Some involve self-discipline (see Thaler (1980)), while others have to do with self-signaling (Prelec and Bodner (2000)). In Section 4 we discuss the relationship between our explanation and these. We also comment on the optimality of decisions governed by mental accounting of the type discussed here. We first turn to discuss overgeneralizations, to what extent they can be rationalized, and why they may be a prevalent mode of reasoning.

2. The Categorical Imperative for Multi-Selves

The notion that people make single decisions as if they were to commit to the same decision over and over again in future decision problems may seem odd at first glance. Why would people overgeneralize their choices in this way? In this section we argue that this mode of reasoning about choice is not completely foreign to the human mind. The next section argues further that it can also be optimal under certain constraints.

A famous example of overgeneralization in reasoning about choice is Kant's categorical imperative. This dictum, namely, that one should behave in a way that one would like everyone to behave, suggests that a decision maker evaluate her choice as if all other relevant

decision makers make the same choice. This is an overgeneralization because different decision makers are sovereign and independent. There is no causal relationship that would justify the generalization as a descriptive theory of what would actually be the outcome of a given choice. Yet, this is a common mode of reasoning.

One reason for making choices in accordance with the categorical imperative is attaching value to following rules as such. Suppose that Sally has to decide whether to pollute the environment. Knowing that her impact on the environment is miniscule, and that other decision makers are independent of her, she knows that others need not follow her decision. Yet, if she finds intrinsic value in behaving in a way that she would like others to behave in, she will make a choice as if she believed that others would follow her choice. This is an *over*-generalization to the extent that Sally knows that it is not supported by a causal relationship. But it may well be Sally's normative criterion for decision making.

Another reason for overgeneralization of choices may be more pragmatic. Consider the celebrated Prisoner's Dilemma (PD) game:

		Player II	
		C	D
Player I	C	3 , 3	0 , 4
	D	4 , 0	1, 1

Whereas D is a dominant strategy for each of the two players, the outcome (D,D) is Pareto dominated by the outcome (C,C). Rational choice dictates that each player know what is and what is not under her control. For instance, since the player II's choice is not under player I's control, and, further, there is no causal relationship between I's choice and that of II's, player I is rightly comparing her payoffs in each column separately, and not along the diagonal.

Indeed, this is the problem exposed by the PD example: in many social situations the structure of the game is similar to the PD game. In these situations, one cannot trust individual optimization to result in a socially desired outcome. Rather, in order to obtain such an outcome, or at least to make it an equilibrium of the game, one has to change the game. One way to do so, and perhaps the most effective one, is to change the payoffs of the players by education, inculcation of guilt, and so forth. Specifically, if all parents in a society condition their children to vividly imagine "what would happen if everybody behaved this

way", these children will, in essence, compare payoffs along the diagonal and prefer C to D. Any two such children who play the game will engage in overgeneralizations: they will behave *as if* the other followed their example, while in reality both are independent and sovereign decision makers. But, mistaken as they are, they will fare better than will two other children who were not trained to overgeneralize.[1]

It follows that, taking a social-evolutionary approach, one can justify overgeneralization as a mode of reasoning about choice in certain situations, such as the PD game. Moreover, casual observation supports the conjecture that people do tend to think in terms of "What would happen if everyone did the same as I do?" Observe that the overgeneralization discussed in Section 1 employs this mode of reasoning, applied to future agents of the *same* decision maker, rather than to different people in a society. For instance, consider Mr. S of Example 1, who wonders whether he should buy the cashmere sweater. Mr. S considers his future decisions, and imagines the "agent" or "self" of himself in each of many decision periods in the future. If he considers these agents, or "selves", as independent and sovereign decision makers, he can go ahead and buy the sweater as if this were a one-shot decision with no further repercussions. But if he imagines that all future selves will behave as he does, he performs an overgeneralization that is very similar to following the categorical imperative, and he may consequently decide to forego the luring purchase.

The overgeneralization of choice across individuals, as suggested by the categorical imperative, may seem less justifiable than the same overgeneralization in the case of multi-selves of a single individual. First, one often tends to believe that different selves of the same individual will have a larger degree of affinity that will different individuals in a society. Selves that occupy the same mind may know more about each other and also care more about each other than do individuals who belong to the same society. Second, different selves of a single individual are called upon to act in a sequential manner, where they can, in principle, know the choices of past agents. Hence future selves can, though they are by no means compelled to, decide to follow the example of earlier selves. By contrast, individuals in a society may often play a simultaneous move game (as the PD above), in which they do not have the information required to adopt such a strategy.

Naturally, the overgeneralization employed by the categorical imperative is inherently ambiguous. For instance, the proposition "John sits on his lawn on Sunday" may be generalized to "Everyone sits on their lawn on Sunday", to "Everyone sits on John's lawn on

[1] To the extent that parents identify with their own children, the decision whether or not to train one's child to overgeneralize generated a PD-like game among the parents. Each parent would prefer that her child would not overgeneralize, irrespective of the others. But all parents will be better off if they all train their children to overgeneralize than if they all train their children not to.

Sunday", to "John sits on his lawn every day", and so forth. John may find the first generalization rather nice, the second – a physical nightmare, whereas the third – physically possible and even pleasurable but morally reproachable. Should John then sit on his lawn on Sunday? The answer depends on the type of generalization that John finds the most salient or compelling. By a similar token, the examples we discussed in Section 1 involve overgeneralizations that are not uniquely defined. Our point is precisely that different contexts give rise to different rankings of generalizations in terms of their salience. Thus we attempt to explain certain phenomena of mental accounting by the categorical imperative applied to multi-selves, in conjunction with the ambiguity of generalizations.

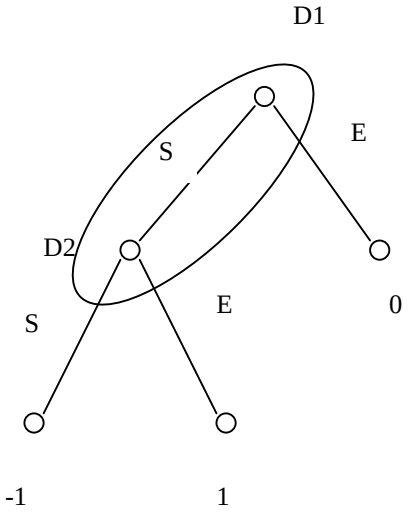
3. Absentmindedness in Game Theory

Despite the similarity of overgeneralizations in cases of mental accounting to those dictated by the categorical imperative, the question remains, why would people make these overgeneralizations? After all, unless one takes a normative view that adopts the categorical imperative axiomatically, it remains erroneous reasoning. Why would people constantly engage in such an irrational mode of decision making?

We suggest here that, should memory fail to be perfect, one may find that overgeneralization is part of a procedure yielding optimal decisions. To see this, consider the example of the Absentminded Driver of Piccione and Rubinstein (1997a).[2] A driver is on a highway. She needs to take the second exit to get to her destination, which would result in a payoff of 1. Taking the first exit yields a payoff of zero, whereas staying on the highway past the second exit yields a payoff of –1.[3] This is a very simple choice problem. There is one twist, however: when the driver gets to the second exit, she will not remember that she has already past one exit. That is, she has imperfect memory. The game may be described by the following tree:

[2] Piccione and Rubinstein (1997a) present the example as a paradox. Several articles and comments in the same volume comment on the example and offer resolutions of the paradox. See Aumann, Hart, and Perry (1997a,b), Battigali (1997), Gilboa (1997), Grove and Halpern (1997), Halpern (1997), Lipman (1997), and Piccione and Rubinstein (1997b).

[3] The payoffs used here are slightly different from those in the original example of Piccione and Rubinstein.



In this game there is a single player, who moves first in decision node D1 and then in decision node D2. In both nodes, she can either exit (E) or stay (S). The payoff is 1 if and only if she exits at node D2.

However, when the driver is at D1 and when she is at D2 she only knows that she is either at D1 or at D2. This fact is reflected in that D1 and D2 belong to the same information set (denoted by the ellipse). It implies that the driver cannot choose a strategy that assigns a different move to the two nodes, since she knows in advance that she will not know which node it is that she will be at.

Let us restrict attention to pure strategies.[4] The optimal strategy, namely to play S at D1 and then to play E at D2, is not feasible at all, as it selects different moves at D1 and at D2. One pure strategy is to select E at both nodes, which results in leaving at the first exit with payoff 0. The second pure strategy is to select S at both nodes, which results in payoff –1. Thus, knowing that it is impossible to choose the second exit, the driver will choose to exit immediately.

Consider now the following example. John likes expensive wine, but he cannot afford to have it everyday. It would be optimal for him to buy a $30 bottle four times a month, and to buy a $10 bottle on the other days. But John is aware of his own limitations. Despite the fact that he can normally count up to 3, he knows that he will not be able always to remember how many times he has bought an expensive bottle that month. He can easily implement a simple rule that says, “Buy a $30 bottle” or one that says, “Buy a $10 bottle”. But implementing a

[4] Piccione and Rubinstein (1997a) observe that in game of this type a mixed strategy may yield a higher payoff than any of the pure strategies in its support.

rule such as, "Buy a $30 bottle if you haven't already bought such bottles four times this month, and otherwise buy a $10 bottle" requires better memory than John's is. Knowing this, John decides always to buy a $10 bottle.

Thus, when John selects a strategy for all his future selves, that is, when he chooses a rule, he is in a similar situation to the absentminded driver above. He knows that any decision he makes at the current node may be repeated in all future nodes, and that he cannot signal to his future selves what is his current choice. If he were to choose instructions for his future selves, these instructions would have to be identical for the different selves. Under this constraint, it is optimal for John always to buy inexpensive wine (exit immediately) than always to buy expensive wine (stay throughout).

We have thus established that overgeneralization is related to the ex-ante optimal strategy, given the constraint of imperfect memory. This simply follows from the assumption that memory does not allow the decision maker to give each future self a possibly different instruction. But it is still not clear that the ex-ante optimal strategy will indeed be followed. When John walks into the store on a particular day, he knows that he can afford the more expensive bottle *just this once*, and actually even four times each month. What, then, would prevent him from disobeying his own rule?

Indeed, the same question is raised by Piccione and Rubinstein (1997a) regarding the absentminded driver when she is at a decision node.[5] Gilboa (1997) argues that the problem should be analyzed as a game among the agents of the driver, and that the problem bears some resemblance to the Prisoner's Dilemma. Specifically, each agent has an incentive to disobey the rule, assuming the other agents obey it, but all agents are better off if they all obey the rule than if they all disobey it. Similarly, John will be better off if his future agents follow the categorical imperative than if they do not, just as a society may be better off if all its agents follow Kant's dictum than if they do not. Thus, behaving as if decisions will be repeated is an optimal strategy ex-ante. But, should each agent make her own choice optimally, this strategy need not be followed. Differently put, for future agents to follow this strategy, they need to exercise some degree of self-discipline.

4. Self-discipline, Optimality, and Self-signaling

Self-discipline We concluded that some notion of self-discipline is required in order to implement the ex-ante optimal strategy in the game of Section 3. It is therefore natural to

[5] The paradox that Piccione and Rubinstein discuss stems from the conflict between the notion of ex-ante optimality, and the interim calculations that the player will conduct while she is playing the game, and taking her strategy as given.

ask, does one need the assumption of imperfect recall? Can we not explain why John buys only $10 bottles by self-discipline alone?

Indeed, one may try to suggest such an explanation. If John had unlimited self-discipline and infallible memory, he could choose four days a month on which to buy a more expensive wine, and follow this strategy with no difficulty. Why doesn't he? One explanation that relies solely on self-discipline would suggest that John can resist the temptation to buy a $30 bottle, but only as long as he hasn't started consuming such bottles. Once he has relished a few, he will not be able to resist buying more. That is, such an explanation would be along the lines of addiction: the expensive bottle can be resisted only before it has been consumed.

The addiction explanation seems counterintuitive in this example. John has probably enjoyed $30 wine bottles in the past. It seems more likely that John cannot restrict himself to four expensive bottles a month because his memory does not provide the technology to implement such a strategy, rather than that he will become addicted to expensive wine.[6]

Self-signaling Some of the phenomena we discuss are closely related to issues of self-signaling as well. (See Ainslie (1992) and Prelec and Bodner (2000)). For instance, the family that drives to the ball game may wish to prove to themselves that they are not the kind of people who spend money on products they do not need. Mr. S may also feel that walking into a store and buying an expensive sweater is not the kind of thing that he does. He may be upset to find out that he, too, and not only his teenage daughter, is a buying type.

Observe that self-signaling explanations also assume generalizations. Specifically, in a self-signaling model one observes one's actions and infers from them what is one's type. But the very notion of a "type" presupposes that there are people who *generally* are prone to certain types of behavior. A model that employs types may provide another rationale for overgeneralizations: the choice in a given problem provides information regarding the decision maker's type, which, in turn, changes the probability of future actions, typically increasing the probability of repeating the same choices. However, in some of the examples above, self-signaling does not seem to be the most intuitive explanation. For instance, John knows that he prefers expensive to inexpensive wine. He chooses to restrain his consumption due to financial constraints alone, not in order to avoid certain realizations about himself.

Optimality John's optimal strategy in the game of Section 3 is optimal ex-ante, but it does not induce an optimal choice for each agent of John's. But some of the mental

[6] There are, however, other ways in which self-restraint can be explained. See, for instance, Asheim (1997), Caillaud, Cohen, and Jullien (1999), and Carrillo (2000).

accounting phenomena discussed here can be viewed as optimal strategies (for the player) that also induce optimal choices for each agent, despite memory limitations. For instance, if John decides to have an expensive bottle once over each weekend, he might be viewed as using the calendar as a memory aid for his wine consumption. He will end up choosing the optimal strategy without memory constraint: he will buy expensive bottles precisely four times a month. Moreover, none of his agents will have an incentive to deviate from this strategy, as it is globally optimal. Similarly, Mr. S of Example 1 may like to buy a luxurious item only once a year. Choosing to do so on his birthday guarantees that he will not exceed this pre-assigned limit, without having to ask himself every day whether or not he has already bought a luxurious item this year. (Of course, he will have to ask himself every day whether it is his birthday.)

In the absence of sufficient memory aids, however, the optimal strategy under the memory constraint will, in general, be sub-optimal. For instance, Mr. S may be able to afford two cashmere sweaters a year. Due to imperfect memory he has to restrict himself to one sweater on his birthday, and to forego the second because he has no way to remember how many sweaters he bought himself as unbirthday presents. The sunk cost effects that make the family of Example 2 drive through a snowstorm to see a game do not lead to an optimal choice, even though they aid in the implementation of an ex-ante optimal strategy. Finally, John might wish to consume an expensive bottle of wine precisely once a week, but not necessarily on a weekend, or not even on a given day of the week. In this case, consuming an expensive bottle on the weekend is still a sub-optimal strategy.[7]

To conclude, imperfect memory may make the optimal choice of a player sub-optimal for some of her future agents. Mental accounting may be a way to implement the ex-ante optimal strategy. Moreover, when mental accounting is sophisticated enough to employ memory aids, it can help implement a strategy that is optimal for a player with perfect memory.

5. Summary

Certain type of phenomena that fall under the broad category of mental accounting can be explained based on the premise that decision makers evaluate specific acts as if they were chosen not only once but repeatedly for a sequence of problems. This is a type of overgeneralization that is reminiscent of the categorical imperative. Indeed, it is formally equivalent to it when one considers the society of selves, or the future agents of an individual. Moreover, if we assume that memory is imperfect, it turns out that not all strategies are

[7] This point is due to Juan Carrillo.

implementable. Only those that are generalized enough may be considered by an individual ex-ante.

References

Ainslie, G. (1992), Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person, Cambridge University Press.

Asheim, G. (1997), ”Individual and Collective Time-Consistency”, Review of Economics Studies, 64, 427- 443.

Aumann, R.J., S. Hart, and M. Perry (1997a), "The Absent-Minded Driver", Games and Economic Behavior, 20, 102-116.

Aumann, R.J., S. Hart, and M. Perry (1997b), "The Forgetful Passenger", Games and Economic Behavior, 20, 117-120.

Battigali, P. (1997), "Dynamic Consistency and Imperfect Recall", Games and Economic Behavior, 20, 31-50.

Caillaud, B., D. Cohen, and B. Jullien (1999), “Towards a Theory of Self-Restraint,” mimeo.

J. D. Carrillo (2000), “Self-Control, Moderate Consumption, and Craving”, mimeo.

Gilboa, I. (1997), "A Comment on the Absent-Minded Driver Paradox", Games and Economic Behavior, 20, 25-30.

Grove, A. and J. Halpern (1997), "On the Expected Value of Games with Absentmindedness", Games and Economic Behavior, 20, 51-95.

Halpern, J. (1997), "On Ambiguities in the Interpretation of Game Trees", Games and Economic Behavior, 20, 66-96.

Kahneman, D. and A. Tversky (1979), “Prospect Theory: An Analysis of Decision Under Risk,” Econometrica, 47: 263-291.

Kahneman, D. and A. Tversky (1984), “Choices, Values, and Frames,” American Psychologist, 39: 341-350.

Lipman, B. (1997), "More Absentmindedness", Games and Economic Behavior, 20, 97-101.

Piccione, M. and Rubinstein, A. (1997a), “On the Interpretation of Decision Problems with Imperfect Recall,” Games and Economic Behavior, 20, 3-24.

Piccione, M. and Rubinstein, A. (1997b), “The Absent-Minded Driver Paradox: Synthesis and Responses,” Games and Economic Behavior, 20, 121-130.

Prelec, D. and O. Bodner (2000), “”, this volume.

Thaler, R. H. (1980), “Toward a Positive Theory of Consumer Choice,” Journal of Economic Behavior and Organization, 1: 39-60.

Thaler, R. H. (1985), “Mental Accounting and Consumer Choice,” Marketing Science, 4: 199-214.

Tversky, A. and D. Kahneman (1981), “The Framing of Decisions and the Psychology of Choice,” Science, 211: 453-458.