June, 1990

Justice and Economic Decision Making: A Model of Akrasia

Robert D. Cooter

Justice and Economic Decision Making: A Model of Akrasia

Robert D. Cooter¹

Economics has provided the most rigorous model of decision making. Unfortunately, its severe rationality assumptions rule out psychological conflict. Modifying the standard model to allow for conflicting preferences creates scope for hesitation, doubt, regret, and akrasia. Akrasia, which is doing wrong knowingly, figures prominently in discussions of morality and justice in classical philosophy. The development of a formal model of akrasia along the lines taken in this essay holds the promise of combining the mathematical rigor of economics, the analytical power of philosophy, and the empirical methods of psychology in the study of justice.

KEY WORDS: justice, decision making, economic, akrasia, Aristotle.

Oscar Wilde defined a "philistine" as someone who knows the price of everything and the value of nothing. Economists believe that the price of everything is its value. I will offer one reason why the standard economic model of decision making excludes values other than price, and I will show how to amend the model to recover some missing values.

The limitation in decision theory can be traced back to an ancient dispute. Socrates argued that a person who really understood the difference between good and evil would choose the good.² He would choose the good

¹Center for the Study of Law and Society, University of California, Berkeley, California 94720.

²The position of Socrates is summarized as follows by I. G. Kidd (1967) in the *Encyclopedia of Philosophy*: "No one voluntarily makes a mistake. By 'voluntarily' he [Socrates] meant consistently with one's true will—that is, to be really happy. Socrates was, of course, familiar with the experience of yielding to temptation, but he explained that in such cases one does not really believe that what one does is bad and does not really know that what is rejected is good. For Socrates 'good' was a term of utility signifying advantage for the doer. Thus, he argued that no one would deliberately choose what will harm him or knowingly reject what will benefit him most. If a wrong choice is made, it must be an intellectual mistake, an error the man who knows could not make." (p. 484)

because doing good is best for himself, given what a person truly is. Injustice, according to this view, results from ignorance. Ignorance might consist in not knowing that an act is unjust, or not knowing that doing justice is always best for the actor.

In contrast, Aristotle recognized that punishment presupposes the possibility of doing wrong knowingly (see Mortimore, 1971; Rorty, 1980), his view is imbedded in familiar features of criminal law. For example, the standard legal criterion for insanity is ignorance of the difference between right and wrong. A person who did not know the difference between right and wrong cannot be lawfully punished for what he did. Furthermore, the more complete the knowledge that an act is wrong, the worse it is to do it. Thus premeditated wrongdoing is punished more harshly than spontaneous wrongdoing.

Doing wrong knowingly was called "akrasia" by Aristotle. The akratic typically yields to an impulse that goes against his considered judgment. Allowing impulse to overpower judgment manifests weakness of will. Wrongdoing can apparently result from deficiencies in either of two distinct faculties, will or reason. Weakness of will causes a person to do wrong knowingly, and ignorance causes a person to do wrong unknowingly.

Over 2000 years after Aristotle's death, fresh essays by philosophers continue to appear on akrasia (e.g., Pears, 1984; Rorty, 1980). Aristotle apparently won the philosophical debate, but akrasia is still excluded from economic decision theory. It is excluded by virtue of severe rationality assumptions. These rationality assumptions provide the most rigorous model of decision making, but they preclude human failures that arise from weakness of the will. I will explain how standard economic decision theory excludes these failures, and I will show how to put them back into the decision calculus by modeling akrasia.

ELEMENTS OF ECONOMIC DECISION THEORY

The economic decision model is based on constrained utility maximization, whose elements I will explain. Let s be a vector that describes the state of the world. Different states correspond to different points in the state-space in Fig. 1. Let the real number line represent a rank ordering from bad to good, as indicated in Fig. 1, with "worse" corresponding to smaller numbers, and "better" corresponding to larger numbers. The decision maker's utility function $u(s)$ is a mapping of states into the real number line, as illustrated in Fig. 1. The mapping assigns larger numbers to more preferred states and smaller numbers to less preferred states.

Fig. 1. Preference over states

The preceding account of utility connects directly to prices. To understand the connection, suppose the initial state of the world subsequently changes and becomes worse from the decision maker's viewpoint. An amount of money usually exists that would perfectly compensate for the change. Compensation is perfect when the decision maker is indifferent between having the money and the worse state of the world, or not having the money and the better state of the world. This amount of money, whose technical name is the "compensating variation," is the amount that the decision maker would be willing to pay for the change. In a perfectly competitive market, the price of a good equals the amount that each consumer is willing to pay for one more unit of it. The compensating variation, or something similar to it, is the concept of value used in standard microeconomics, such as cost-benefit analysis. (There are in fact several closely related forms of variation that can be used to measure value in cost-benefit analysis, e.g., the "equivalent variation" rather than the "compensating variation." See Hicks, 1968, pp. 333-335.)

The utility functions describes the decision maker's subjective preferences. Choice is constrained by objective considerations. For example, a consumer is constrained by his budget, and a producer is constrained by the production technology. Let the function $f(s) \leq 0$ describe the objective constraints upon choices. Choices satisfying the constraint $f(s) \leq 0$ are feasible. In the standard economic model, the decision maker maximizes utility subject to the feasibility constraint:

$$\max_s u(s) \text{ subject to } f(s) \leq 0$$

I must relate the standard microeconomic model of decision making to the faculties of mind that cause wrongdoing. Following Plato's *Republic* an analogy can be drawn between the faculties ascribed to decision maker and the parts of the body. The decision maker's preferences especially arise from desires or drives, which Plato associated with the stomach and genitals. Beliefs about cause and effects, including the feasible set, belong to the faculty of reason, which Plato associated with the head. Preferences and beliefs are the subjective elements in the standard economic model of decision making. In Plato's terminology, the two faculties of mind used by the decision maker in the standard economic model are desire and reason.

When decision making is reduced to combining desires and beliefs some elements of real decision making are submerged. Hesitation, doubt, psychological conflict, regret, and remorse are minimized or eliminated. They have little or no place in the standard decision calculus. As a result the standard economic calculus yields a decision but does not model the process of deciding.

In contrast, the doubts, hesitations, and conflicts that arise in the process of deciding figure prominently in psychology and philosophy. These phenomena are aspects of willpower. Plato understood "will" as spiritedness, as in the phrase "a spirited horse." He associated spiritedness with the heart. Following Plato's analogy, I could say that the economic model ascribes head and stomach to the decision maker, but not heart. The aim of this essay is to put heart back into the model.

A MODEL OF PRUDENCE

The simplest way to do so is by modeling the tension between immediate impulse and future preferences. The ability to give the proper weight to present and future is the virtue of prudence. I model prudence and show how weakness of will can result in imprudence.

Fig. 2. Consume or invest.

Figure 2 depicts a choice between consuming \$1 now, or investing and receiving benefits b at the end of one period of time. In the standard economic framework, the decision maker applies his subjective discount rate r to the benefit b in order to reduce it to its present value:

$$\text{present value} = b/r$$

To illustrate, b might equal 1.00 and r might 1.05, which implies a present value of 0.95. The standard decision rule is to invest whenever the present value of investment is larger than the value of immediate consumption. When the alternative is to consume one unit or invest it, this decision rule can be written:

$$\max(1, b/r)$$

The subjective discount rate r of the decision maker may depend upon transitory factors like mood, emotion, or deprivation. To capture this idea assume that r is drawn from a probability distribution $p(r)$, as depicted in Fig. 3. The value of r drawn just before making the decision is denoted r_0 . In the standard model, the decision maker solves

$$\max(1, b/r_0)$$

Whether the solution to this problem results in consuming or investing depends upon the size of r_0 . For large values of r_0 , the solution to this problem favors consuming, and for small values of r_0 the solution favors investing. Consequently, some value of r_0 exists which is a tipping point between the impulse to consume and invest, which is denoted by r^* . (The tipping point is defined by $1 = b/r^*$.) For values of r less than r^* , the decision maker's immediate impulses favor consumption. For values of greater than r^* , the decision maker's immediate impulses favor investment.

Fig. 3. Probabilistic discount.

Fig. 4. Tipping value.

Fig. 5. Conflict.

Assume that investment yields a very large return, so b is substantially greater than 1. Consequently, the discount rate r must be very large in order to reduce b to a present value of less than 1. The tipping value r^* is thus in the right tail of the probability distribution $p(r)$, as illustrated in Fig. 4. Most of the probability density in Fig. 4 lies below r^* . So the decision maker usually draws preferences that favor investing. Sometimes, however, the decision maker draws a value of r_0 from the shaded area in the right tail of the distribution, where r_0 exceeds r^* . In these circumstances, the decision maker's immediate impulse favors consuming rather than investing.

If the decision maker invests, time elapses before realizing the benefit b . During this interval, the mood, emotion, or deprivation level may change, thus causing a change in the discount rate. To capture this idea, assume that the decision maker draws from the probability distribution $p(r)$ a second time, as depicted in Fig. 5. The second value of r is denoted r_1 . At the point of deciding whether to consume or invest, the decision maker has drawn r_0 but not r_1 . Looking ahead, the decision maker knows that r_1 may diverge from r_0 , which creates a potential conflict.

This is a conflict between immediate impulse and the regret that it is likely to cause. The decision maker can be said to regret an action if he

would not have done it given his present preferences. Looking ahead, the decision maker can foresee that he is likely to have preferences in the future that will cause him to regret the decision to consume rather than invest. [A precise mathematical formulation can be given for his "expected regret." The decision maker calculates the expected present value of the benefit from acting, as defined by

$$E(b/r_1) = \int (b/r_1)p(r_1) dr_1$$

When the value $1 - E(b/r_1)$ is negative, the decision maker expects to regret consuming rather than investing. Thus the expected amount of regret equals $E(b/r_1) - 1$.]

A conflict exists between the immediate impulse to consume and the considered preference for investment whenever r_0 exceeds r^* . A considered choice balances the satisfaction of immediate impulse and the expected regret that it causes. This conflict between two alternatives, both of which are desired, can be analyzed with the standard economic tools, but I do not pursue further mathematical elaborations here.³ Rather I consider the interpretation of this conflict.

Prudence consists in acting upon considered preferences rather than immediate impulses. In this model, an imprudent person knows his considered preferences, but his will is not strong enough to act on them. In prudence is a form of akrasia. This model of akrasia can be elaborated formally by proving various theorems, as I have done elsewhere (Cooter 1991). For example, an increase in the variance of $p(r)$ without any change in its mean ("mean preserving spread") usually increases the frequency of imprudent acts and cannot decrease their frequency. However, this essay is not the place to prove theorems. Instead, I will discuss the extension of the model to distributive justice and show some implications of the model for public policy and social theory.

EXTENSION TO DISTRIBUTIVE JUSTICE

Imprudence concerns the relative weight given to the present and future benefits of a single person. In effect, a prudent person gives appropriate consideration to his present and future selves. Distributive justice

³Specifically, a decision function can be written with impulse satisfaction and expected regret as its arguments: $w = w(b/r_0, E(b/r_1))$. The decision function is, in effect, a second-order utility function that describes how the decision maker trades off two motives. A start is made in developing this idea in Cooter (1991).

concerns the relative weight given to the benefits of different people. A just person considers himself and others, and gives each his due. Thus there is a formal similarity between future selves and other people. I will exploit this similarity in order to adapt the model of imprudence to distributive justice.

Suppose the choice in Fig. 2 is between benefiting myself and benefiting others. The self-regarding act, previously called "consume," benefits me by 1 unit. The other-regarding act, previously called "invest," benefits others by " b " units. The decision maker decides what to do by comparing the size of the benefits. The comparison is modeled as a discount rate r , which depends upon mood, emotion, deprivation, and other transitory phenomena. Just before the decision is made, the decision maker draws r_0 from $p(r)$, as depicted in Fig. 3: b/r_0 measures the decision maker's spontaneous impulse towards generosity. If he acted purely on impulse, he would solve $\max(1, b/r_0)$.

The tipping point between self-regarding and other-regarding acts is the value of r , denoted r^* , which is defined by $1 = b/r^*$. If r_0 is less than r^* , the decision maker's impulse is to do the other-regarding act. If r_0 is more than r^* , the decision maker's impulse is to do the self-regarding act.

Suppose that b is so much larger than 1 that justice requires doing the other-regarding act. If r^* is the right tail of the distribution $p(r)$ as depicted in Fig. 4, the person usually has the impulse to act justly, but he sometimes has the impulse to act unjustly.

Rather than acting on impulse, the decision maker may act on his considered judgment. Part of a considered judgment is comparing the strength of the impulse to act with the regret that it causes. This balance can be represented in mathematical notation. The strength of the impulse is indicated by the amount that the value of the self-regarding act exceeds the discounted value of the other-regarding act, using the spontaneous discount rate r_0 . In notation, the strength of the impulse is measured by $1 - b/r_0$.

To compute regret, the decision maker considers that he will draw r from $p(r)$ again, as depicted in Fig. 5. At a future date, the other-regarding decision will be valued at b/r_1 , instead of b/r_0 . The difference in value between the other-regarding and self-regarding acts, denoted $b/r_1 - 1$, measures the strength of the future regret that he will feel. The expected regret can be written $E(b/r_1) - 1$. The decision maker thus decides what to do by comparing the impulse $1 - b/r_0$ and the expected regret $E(b/r_1) - 1$.

In previous models economists have explored the consequences of decision makers with preferences over the utility or income enjoyed by others. (For example, there is a large literature on "Pareto optimal income distribution" beginning with Hochman and Rodgers, 1969.) The formal innovation in my model is to allow these preferences to vary over time. Vari-

ation over time creates the possibility of a conflict between immediate impulse and considered preferences, as required for akrasia.

RATIONALE FOR PATERNALISM

The model of akrasia provides a rationale for paternalistic policies of income redistribution. For example, some people are so imprudent in youth that they save nothing. In old age they will regret this imprudence. However, majority rule averages over the votes of young and old. If people are prudent on average, a majority may vote for a social security system that imposes compulsory savings on young people. A social security system helps to secure a just distribution of resources between old and young by overriding imprudent impulses.

The argument for paternalism can be formulated rigorously with the help of the model. Suppose that different people have different preferences r at any point in time, but everyone draws their preferences from the same distribution $p(r)$ at each point in time. Thus the average preferences of voters correspond to the expected preference $E(b/r)$ of a particular individual. The winner under majority rule is the alternative that the median voter most prefers. Under these assumptions, the average preferences of voters at any point in time may align with the average preferences of an individual whose preferences vary over time. [This result is achieved when $p(r)$ is symmetrical, so that the median is also the mean.] When this condition is met, the collective choice of voters corresponds to the considered choice of each individual as defined above. Thus the collectivity justifies paternalism on the grounds that it satisfies the considered preferences of every individual.

MORAL EDUCATION

An objective of moral education is to help the individual to master his impulses. Stability comes from habit, and habit is built up by repetition. For example, a person who acts prudently in circumstances that tempt him to act impulsively may thereby strengthen his will. Repetition of right behavior, when combined with an understanding of why the behavior is right, may result in the internalization of norms.

This argument can be clarified with the help of the model. More stable behavior is represented in Fig. 4 by reducing the variance in the probability distribution $p(r_0)$. Thus usual justification for moral education is that the variance of $p(r)$ diminishes with repetition of the right behavior.

A rational decision maker will take steps to strengthen considered preferences so that they will prevail over immediate impulses. One way to tip the scales is by making a commitment not to act on impulse. A commitment is an act that greatly raises the cost of doing the wrong thing. In terms of the preceding figures, a commitment increases the cost of "consuming" or doing the "self-regarding" act. After commitment, the benefit of the wrong act falls from 1 to 1 minus the commitment's value.

To illustrate, suppose a physician with a cocaine addiction enters a drug rehabilitation program. The physician agrees to submit to weekly urine tests to determine whether he has remained free from cocaine. To increase his incentive to keep clean, he drafts a letter to his local medical society confessing to his cocaine addiction. He gives the letter to the testing agency, which promises to mail it if, and only if, he tests positive for cocaine use. The physician's commitment greatly increases the cost to him of wrongdoing.

This is one of the many clever examples of real-life commitments catalogued in Schelling's (1975, 1984a, 1984b, 1986) pioneering research. In general, the ways in which a person commits in order to strengthen his best impulses are similar to the ways in which the principal creates incentives for an economic agent. For example, a shipper of cargo may promise to pay stipulated penalties for late delivery caused by his own negligence. Commitment is similar to inserting penalties in a contract that are triggered when the principal detects a proscribed act by the agent.

CONCLUSION

Ignorance and weakness of will are two causes of injustice. These two distinct causes of injustice have different remedies. Ignorance can be cured by communication and education. Overcoming weakness of will requires discipline and commitment. Elaborating the remedies is a task for social science. Economics has provided social science with its most rigorous model of decision making. This model has been carefully crafted and adapted to cope with limited computational powers and costly information. For example, the economic analysis of torts and crimes has explored the question of whether ignorance should be a defence against guilt or liability.

However, this model has not been crafted or adapted to cope with weakness of will. The severe rationality assumptions of the standard model rule out psychological conflict. Modifying the standard model to allow for psychological conflict creates scope for akrasia, which is central to morality and justice. The development of a formal model of akrasia along the lines taken in this essay holds the promise of combining the mathematical rigor

of economics, the analytical power of philosophy, and the empirical methods of psychology in the study of justice.

REFERENCES

- Cooter, R. D. (1991). Lapses, commitment, and akrasia in torts and crimes: Towards an economic theory of the will. *Int. Rev. Econ.* (in press).
- Hicks, J. (1968). *Value and Capital*.
- Hochman, H., and Rogers, J. D. (1969). Pareto optimal redistribution. *Am. Econ. Rev.* 59: 542-557.
- Kidd, I. G. (1967). Socrates. In *Encyclopedia of Philosophy*, pp. 480-486.
- Mortimore, G. (ed.). (1971). *Weakness of Will*.
- Pears, D. F. (1984). *Motivated Irrationality*.
- Rorty, A. (1980). Akrasia and pleasure: Nicomachean Ethics Book 7. In *Essays on Aristotle's Ethics*.
- Schelling, T. (1978). Egonomics, or the art of self-management. *AER Papers Proc.* 68: 290-294.
- Schelling, T. (1984a). *Choice and consequence*.
- Schelling, T. (1984b). Self-command in practice, in policy, and in a theory of rational choice. *AER Papers Proc.* 74: 1-11.
- Schelling, T. (1986). Against backsliding. In Foxley, A., McPherson, M. S., and O'Donnell, G. (eds.),