

Treating Yourself Instrumentally: Internalization, Rationality and Law

by

Robert Cooter

Herman Selvin Professor of Law, University of California at Berkeley
Bacon-Kilkenny Visiting Professor, Fordham Law School, spring 2003.

Revised June 2003

“Robert Cooter” <rdc@mail.law.berkeley.edu>

Abstract:

Economics, which has greatly advanced deterrence theory, has made no contribution to understanding how law changes peoples' values. I combine economics and psychology to offer the framework for such a theory. Rational self-development involves treating yourself as the means to achieve your ends. Pursuing ends changes objective opportunities and subjective values. The theory of cognitive dissonance predicts how changing opportunities changes values. If the connection between opportunities and values is strong enough, one choice may dominant the alternatives. A dominant choice is preferred when evaluated by using any likely values. I call a choice that is better with respect to the likely values of a decision-maker a “Pareto self-improvement.” By creating opportunities for Pareto self-improvement, law changes peoples' values. Law's ability to change behavior by changing values, however, occurs slowly, whereas law changes behavior quickly by other means such as deterrence.

Treating Yourself Instrumentally: Internalization and Rationality

“The Puritans came to Massachusetts seeking a life of greater restrictions.”
–Garrison Keeler.¹

Effective law appeals to good people and deters bad people. Deterrence theory has improved greatly in recent years through the application of economics and psychology. In contrast, the theory of internalized values has improved modestly, partly because of the failure of economics to progress in this area. Economists typically assume that a person pursues his self-interest as he perceives it. Perceived self-interest presupposes personal goals, including the central values by which a person defines himself. Instead of explaining how people acquire their goals, economics conventionally “takes preferences as given”. Economics thus offers no account of how a person becomes the self in which he is interested.

To solve this problem, economics needs a theory of self-development, including a theory of endogenous preferences. The fluorescence of behavioral economics in the 1990s brought economics into intimate contact with cognitive psychology. The contact, however, involves only a small part of psychology. Whereas cognitive psychology concerns beliefs, much of the rest of psychology concerns values and motives. A more complete interaction between economics and psychology, that brings economics in touch with values and motives, will contribute to the development of an economic theory of endogenous preferences.

* Herman Selvin Professor of Law, University of California at Berkeley.

I will sketch a way to make the connection. A moral principle holds that we should treat others as ends. This paper concerns the opposite problem: treating ourselves as means. We treat ourselves as means when we shape ourselves to achieve our ends. Rational self-development involves commitment to a discipline that changes a person's skills and values. As Garrison Keeler's droll quotation on the New England Puritans suggests, we choose the constraints that change us.

To illustrate the problem of self-development, consider a young person with talent for music and accounting who must choose between conservatory and business school. Business school and conservatory will impart different skills. Accountants and musicians have different opportunities to make money and music. Business school and conservatory will also impart different values. Accountants and musicians disagree systematically about the importance of money and music for a good life. Thus the student must make choices affecting both skills and values.

To analyze this problem, I combine the economic theory of decision-making and the psychological theory of cognitive dissonance. Economic theory characterizes preferences and opportunities, and psychological theory shows their connection. After sketching the foundations for a theory of rational self-development, the conclusion briefly discusses some potential applications to law.

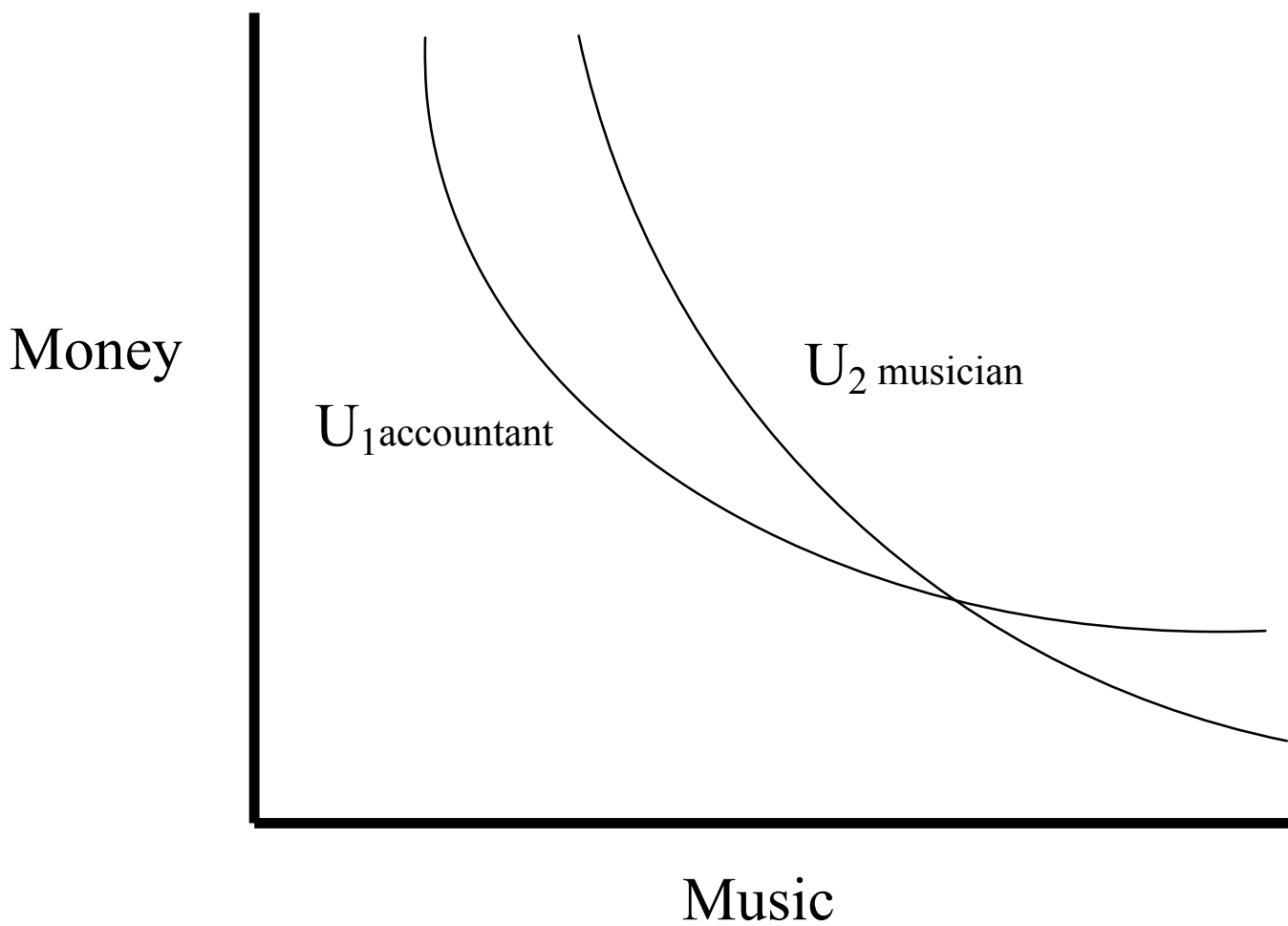
I. Example: Accountant or Musician?

I will analyze the expected changes in young person choosing between conservatory and business school. Figure 1 depicts the expected change in

¹ I heard Garrison Keeler say this on a radio broadcast of his show, "The Prairie Home Companion."

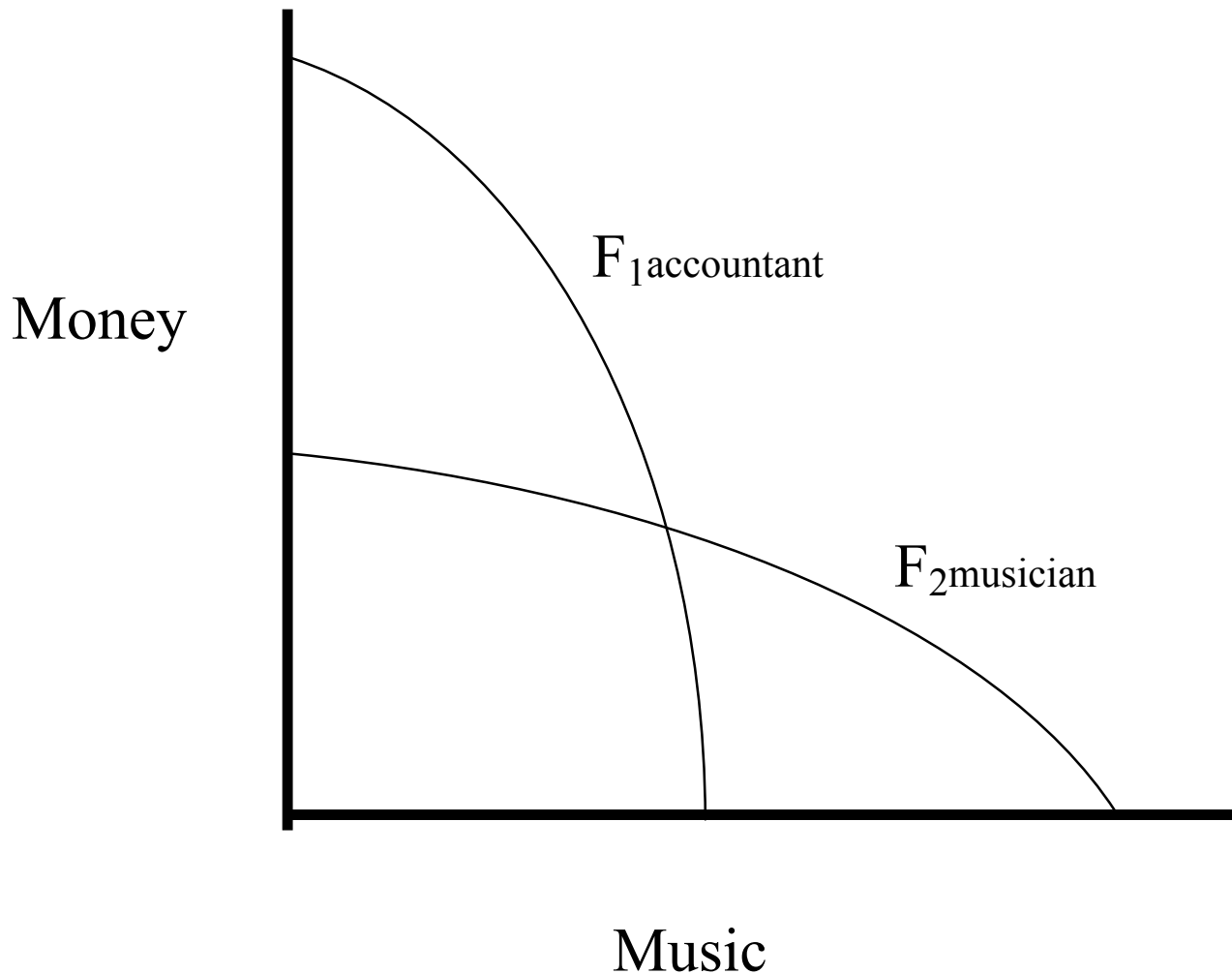
values. Business school will give him a keen appreciation of wealth, as indicated by $U_{1\text{accountant}}$. Conservatory will give him a keen appreciation of esthetic values, as indicated by $U_{2\text{music}}$. To appreciate the difference in values, consider the slope of the two utility curves at the point where they intersect. His expected preferences as a future musician indicate that he would trade a lot of money for more music, and his expected preferences as a future accountant indicate that he would trade a lot of music for more money.

Figure 1: Future Preferences of Accountant and Musician



Having characterized preferences in Figure 1, I characterize opportunities in Figure 2. A skillful accountant with a personality for business has different opportunities than a skillful musician with a personality for performing. Specifically, an accountant has more opportunities to make money and a musician has more opportunities to make music. Figures 2 depict this difference, where the accountant has opportunities $F_{1 \text{ accountant}}$ and the musician has opportunities $F_{2 \text{ musician}}$.

Figure 2: Opportunities of Musician and Accountant



II. Cognitive Dissonance and Market Signaling

Preferences in Figure 1 represent subjective trade-offs, and opportunities in Figure 2 represent objective trade-offs. Now I turn to their psychological connection in the theory of cognitive dissonance.² According to this theory, commitment to an activity that changes opportunities as in Figure 2 will cause a change in values as in Figure 1. Thus cognitive dissonance theory asserts that an actor who changes his objective trade-offs will cause his subjective trade-offs to change.

To see why, I briefly relate commitment in economics and dissonance in psychology. In economics, “commitment” especially refers to a decision that raises prices to the decision maker. To illustrate from Sun Tzu’s The Art of War, an army may commit to advance by burning the bridges behind it, thus raising the price of retreat. In notation, an actor commits to x by raising the relative price of not- x . Specifically, let x denote “advance,” and let not- x denote “retreat.” The army commits to x by burning the bridges it has crossed, which raises the price of not- x .

Note that raising the relative price of not- x is materially equivalent to lowering the relative price of x . Thus we could also say that an actor commits to doing x by lowering its price relative to the price of not- x .

² The theory of cognitive dissonance was especially developed by L. Festinger. See Festinger, L. (1962). A theory of cognitive dissonance. Stanford, Calif., Stanford University Press. A good exposition is in Griffin, E. A. (1997). “Chapter 16 Cognitive Dissonance Theory of Leon Festinger,” A first look at communication theory. New York, McGraw-Hill: xxiv, 536, and also in Mills, J. and E. Harmon-Jones (1999). Cognitive dissonance : progress on a pivotal theory in social psychology. Washington, DC, American Psychological Association. For an economist writing on dissonance and morality, see Rabin, M. (1994). “Cognitive Dissonance and Social Change.” Journal of Economic Behavior & Organization 23: 177-194. For early writing by an economist on the subject, see Akerlof, G. A. and W. Dickens (1982).

Why commit? Raising the price of retreat proves to the enemy that the advance is no bluff. Once the invading army must advance, retreat may be the defending army's best move. In general, an actor in a strategic interaction sometimes gains an advantage by raising his own costs of doing something. My concern here, however, is not with strategy but with values.

Whereas economics is clear about how commitment affects strategy, psychology is clear about how commitment affects values. Cognitive dissonance theory predicts that committing to x makes the actor willing to pay relatively more for it. In other words, cognitive dissonance predicts that an act increasing the *objective* price of not-x causes an increase in the actor's subjective price of x. For example, studying music rather than accounting makes getting money hard relative to playing music, which causes the actor to increase his valuation of music relative to money.

Why does the actor's commitment to something increase its subjective value? Most people want to think of themselves as having good judgment. Good judgment implies good outcomes of choices. Thus people want to see their world as obeying this syllogism:

I make good choices.
I chose this.
Therefore this is good.

"Cognitive dissonance" is the psychological term for the discomfort produced when my own choice produce bad outcomes.

The actor who makes a bad choice must live with the discomfort or else change his evaluation of the outcome. Studies by psychologists and sociologists

"The Economic Consequences of Cognitive Dissonance." [American Economic Review](#).

demonstrate the surprising frequency with which people respond to bad outcomes by changing their beliefs about what counts as bad. An outcome regarded as bad by subjects before choosing may seem good after it results from their choice.

To illustrate cognitive dissonance in self-development, consider the student who hesitates between studying accounting and music. Regardless of which alternative the student chooses, he will probably strain to like it. Assume the student chooses conservatory over business school. The student might subsequently seek new friends among musicians rather than businessmen, focus his thoughts on the advantages of music over business, join discussions about the superficial values of the bourgeoisie, and listen to music at dinner rather than the stock market report. Even if the student struggles in conservatory and gets bad results, he may convince himself that his failures are actually good for him.

Now I show how to depict cognitive dissonance graphically. Figures 1 and 2 represent the difference in preferences and opportunities expected by a graduate of conservatory and business school. Cognitive dissonance theory predicts that commitment to a course of study that produces the opportunities in Figure 2 will cause the values in Figure 1. Specifically, the student who commits to business school raises the cost of music relative to money and faces opportunities F_1 , which causes him to acquire preferences U_1 . Similarly, the student who commits to conservatory raises the cost of money relative to music and faces opportunities F_2 , which causes him to acquire preferences U_2 .

This psychological theory connects preferences and opportunities. In economics, the theory of market signaling also connects preferences and opportunities. However, the direction of causation is reversed in the two theories. Whereas cognitive dissonance postulates that opportunities change preferences, market signaling postulates that preferences change opportunities, as I will explain.

In markets where actors rely on signals, changing values can change a person's opportunities. To illustrate, a business prefers to hire a young executive with a keen desire for wealth, whereas an orchestra prefers to hire a young musician dedicated to art. Knowing this fact, a person applying for a job as an executive will represent himself as dedicated to wealth, whereas a person applying for a job as a musician will represent himself as dedicated to art. Many of the signs that indicate motivation by a particular goal are easier to acquire by people who actually have the motivation than by people who lack it. For example, a person who is dedicated to art will spend his leisure time differently than a person dedicated to wealth. Under certain circumstances, how a person spends his leisure time can signal his motivations.

In general, the ability to signal and detect character is one of the most important skills in social life, as well as in experimental games that involves cooperation.³ Having noted the connection between cognitive dissonance and market signaling, I will not explore the latter any further in this paper. (Melvin

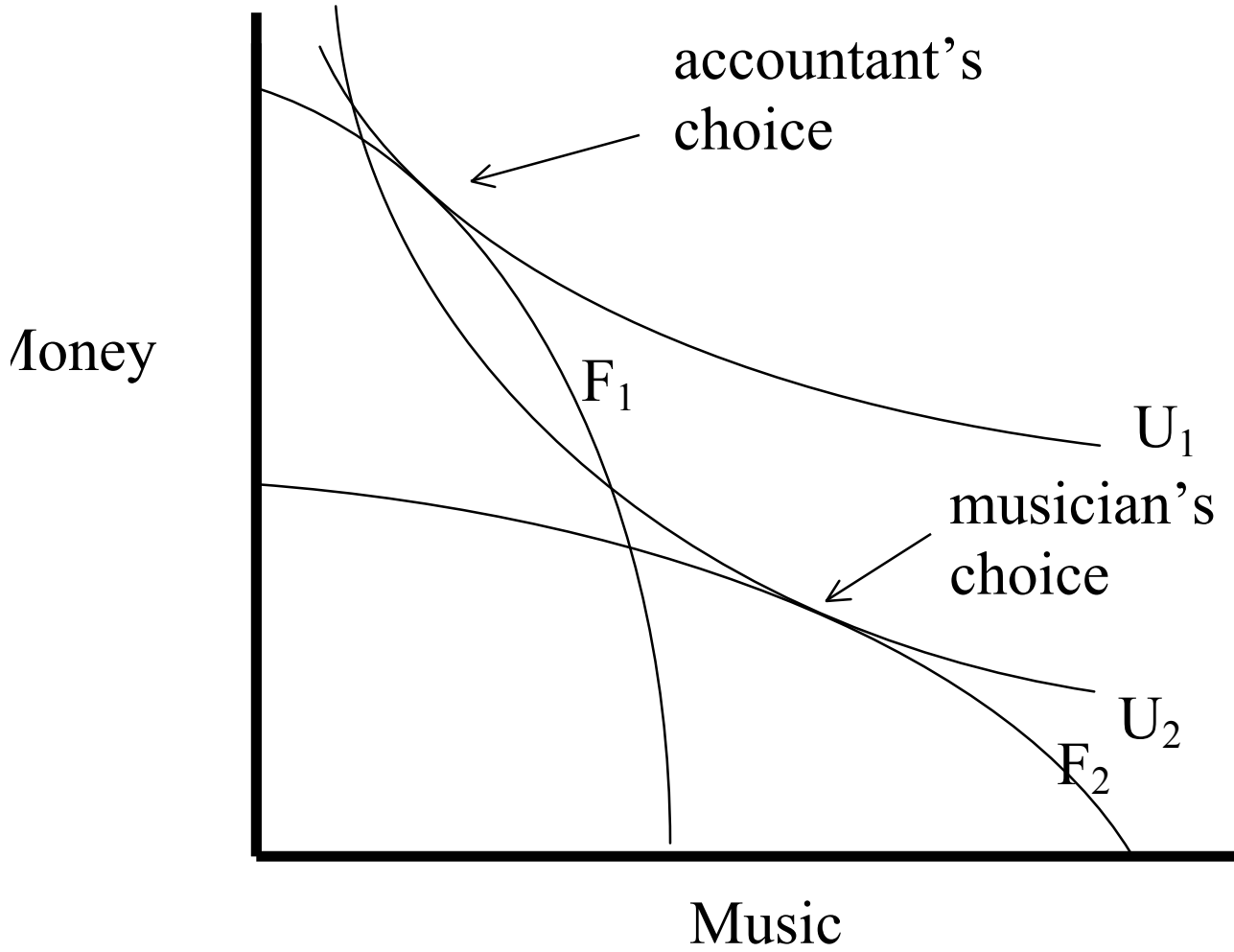
³ Robert Cooter and Melvin Eisenberg, "Fairness, Character, and Efficiency in Firms," 149 University of Pennsylvania Law Review 1717-1733 (2001).

Eisenberg and I have already written on the connection between character and business opportunity.)

III. Meta-Values

Choice occurs when preferences meet opportunities. Figure 3 combines Figures 1 and 2 in order to depict the difference choices that the student will make, depending on whether he becomes an accountant or musician. As an accountant, his preferred point in Figure 3 involves more money than music, and as a musician his preferred point involves more music than money. The problem addressed in the next section is how to choose between going to business school and reaching the accountant's preferred point in Figure 3, or going to conservatory and reaching the musician's preferred point in Figure 3.

Figure 3: Choices by Accountant and Musician



A rational person who understands psychology recognizes that his commitments will change his values. In particular, the student should foresee that his commitment to accounting or music would result in the outcomes depicted in Figure 3. To decide which commitment to make, however, he needs to choose among values. I will suggest how to choose, which involves going beyond conventional economics.

Perhaps the student views money and music as means to something else. For example, the student may want happiness, pleasure, social status, self-fulfillment, moral goodness, or some other general value. Connecting the choice of the particular alternative to meta-values can solve the student's problem.

To illustrate, assume that the student wants to be happy above all. The student might think of happiness as a meta-value or meta-preference that lies behind his desire for money and music. He recognizes that studying music will cause him to gain relatively more happiness from music than money, and studying accounting will cause him to gain relatively more happiness from money than music. If he could predict with confidence how much happiness each one will bring him, then he might choose the career that promised more happiness.

The preceding discussion assumed that happiness is the meta-value that the student maximizes. Perhaps some people seek happiness and others seek

pleasure, social status, self-fulfillment, moral goodness, etc. In this paper I will not discuss whether everyone seeks, or ought to seek, happiness. Instead, I want to focus on the fact that young people seldom know what they really want to accomplish or who they really want to be. In my example, the student is probably uncertain about whether he wants happiness more than, say, fame. And even if he is certain that he wants happiness more than fame, he is probably uncertain about which career path will bring him more of it. In general, self-development involves self-discovery because the effects of choices on meta-values are uncertain.

The problem of uncertainty raises technical questions about how to extend expected utility theory to choices that affect preferences, but I will not pursue these technical questions here. Instead, I turn to a way of choosing that does not involve the calculus of expected utilities.

IV. Pareto Self-Improvement

Economic theorists have intensively studied choice when the decision maker is uncertain about the external world and certain about his values. Economic theorists, however, have devoted much less analysis to decisions when values are uncertain. Although little economic analysis concerns uncertainty about value, something useful can be learned from what has been done. The problem of uncertain values especially arises when economists advise policy makers who face unfamiliar tradeoffs. Here is a typical example from decision theory: Assume that a public hospital must choose between adopting two procedures, one of which causes injury with high probability and no

deaths, and the other causes death with low probability and no injuries.

Depending on its choice, the hospital can anticipate many injuries or few deaths.

Most hospital officials are uncertain their relative valuation of statistical injuries and deaths. (Many similar examples exist.⁴)

A decision maker may be certain about the relative value of extreme alternatives and uncertain about the relative value of close alternatives. To illustrate, the hospital administrator in this case may be confident that statistical injuries above a certain level exceed the cost of a statistical death (say 50:1 or above), and statistical injuries below a certain level fall short of the cost of a statistical death (say 10:1 or below). In between these extremes, however, the administrator may be uncertain about relative values.

An important task for decision theory is to identify the critical numbers on which such a decision turns. This approach is sometimes called “sensitivity analysis,” because it aims for the minimum information about values to which the actual choice is sensitive. To illustrate, the hospital’s choice is insensitive to a tradeoff of statistical injuries to death that exceeds 50:1 or falls short of 10:1. If the actual choice is sufficiently extreme to fall in the insensitive range, the hospital officials can choose without inquiring further into their values. Conversely, if the actual choice is inside the sensitive range, the hospital officials cannot choose without inquiring further into their values.

⁴ Another example is a choice between two alternative locations for an airport, when the decision maker is uncertain about the relative value of saving time to travelers and reducing noise to residents. Yet another example is a decision about whether to build an atomic energy plant when the decision maker is uncertain about the relative value of electricity and the risk of a nuclear accident.

The hospital's alternatives are lumpy like watermelons, not continuous like gasoline. In general, when choices are lumpy, the best alternative may be so much better than the next best alternative that the choice is insensitive to a range of reasonable values.

Now I will adapt an economic concept in order to perform a sensitivity analysis of the student's choice between accounting and music. The concept of Pareto efficiency provides the basis for a sensitivity analysis of the student's decision. When comparing one alternative to another in conventional welfare economics, a choice is "Pareto superior" if everyone prefers it to the alternative. I have extended this concept to a situation where a person's values can change. When comparing one alternative to another, I describe a choice as "Pareto superior" if it is preferable to the alternative when evaluated using each of the decision maker's possible preferences. In my extension, possible preferences of a single person play the same role as actual preferences of different people in conventional welfare economics. The conventional analysis thus concerns *inter-personal* Pareto superiority and my unconventional analysis concerns *intra-personal* Pareto superiority."

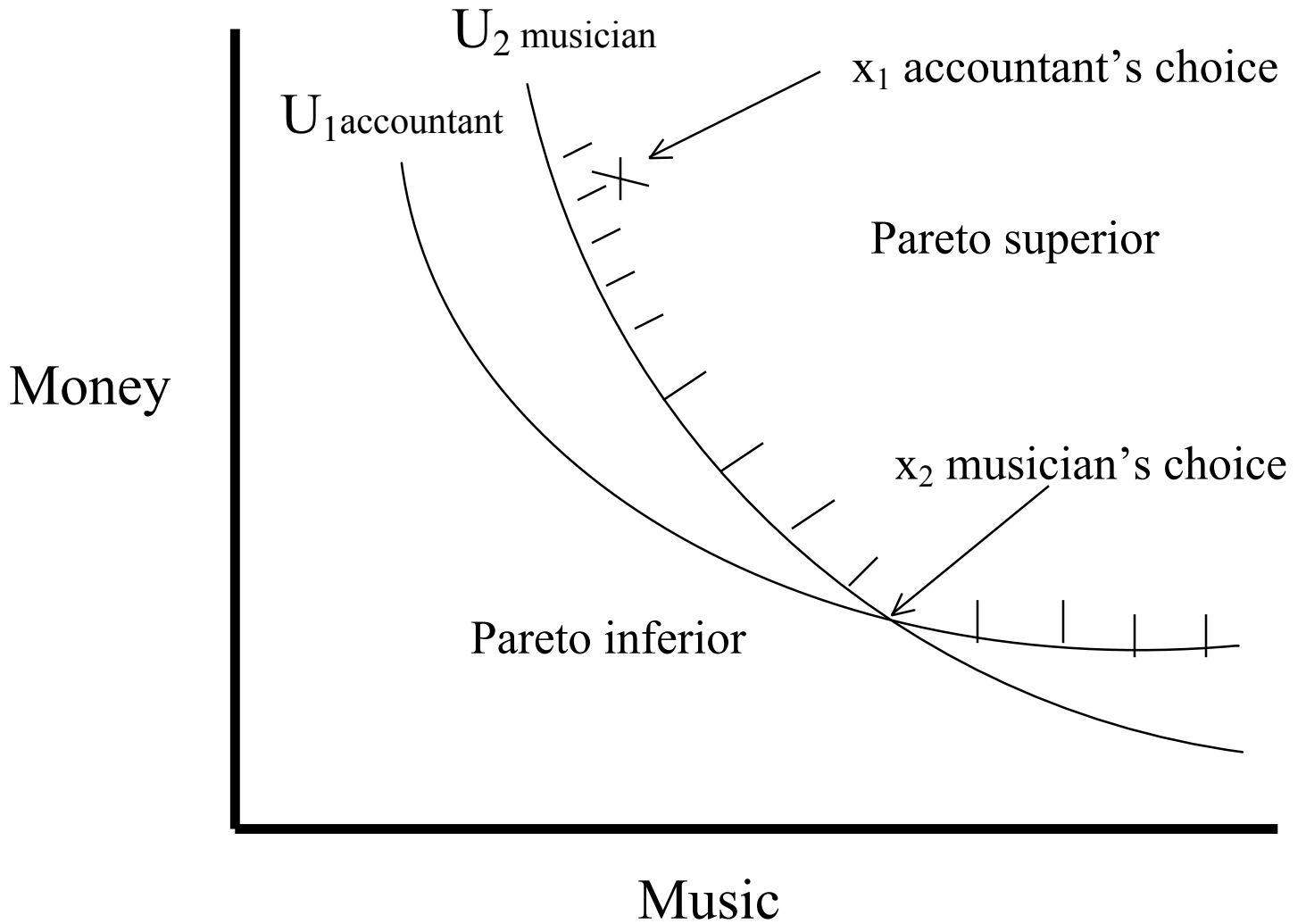
Figure 3 illustrates intra-personal Pareto superiority. Recall that the point labeled "musician's choice" indicates the combination of money and music that the student will choose if he becomes a musician. Also recall that the point labeled "accountant's choice" indicates the combination of money and music that the student will choose if he becomes an accountant. Notice that the accountant's utility curve U_1 in Figure 3 passes through the accountant's

preferred point and above the musician's preferred point. This fact indicates that the accountant likes what he receives better than what he would get as a musician. Also notice that the musician's utility curve U_2 in Figure 3 passes through the musician's preferred point and below the accountant's preferred point. This fact indicates that the musician likes what he would get as an accountant better than what he actually receives as a musician. Thus Figure 3 depicts a situation where the student prefers the combination of money and music that he expects to receive as an accountant rather than a musician, regardless of whether he compares the outcomes with the values of an accountant or a musician. For this student in these circumstances, becoming an accountant is intra-personally Pareto superior to becoming a musician.

Notation helps to clarify the concept of intra-personal Pareto superiority. Let x_1 and x_2 denote two possible outcomes, such as the accountant's choice and the musician's choice in Figure 3. Let U_1 and U_2 represent two possible preference orderings. x_1 is intra-personally superior to x_2 if $U_1(x_1) > U_1(x_2)$ and $U_2(x_1) > U_2(x_2)$.

Figure 3 depicts one choice that is intra-personally Pareto superior to the other. Figure 4 depicts all such choices. Specifically, the set of points labeled "Pareto superior" in Figure 4 contains all combinations of money and music that yield higher satisfaction than the musician's most preferred point, regardless of whether the actor develops musician's preferences or accountant's preferences.

Figure 4: Pareto Self-Improvement



If one alternative is intra-personally Pareto superior to the other, this fact provides a strong reason to choose without inquiring any further into values. To

see why, return to the example of the student. To make a decision, the student would like to meta-values to the alternatives, but the student is uncertain about his meta-values. Assume that the accountant and musician's preferences, U_1 and U_2 , are two extreme possibilities for the rate at which money and music tradeoff in the student's true system of meta-values W . Under this assumption, the student who chooses the Pareto superior alternative necessarily makes a choice that is better by his meta-values. Consequently, the student who grows in self-knowledge will not regret his choice.⁵

To illustrate in notation, I assume that the accountant's and musician's preferences, U_1 and U_2 , are two extreme possibilities for the rate at which money and music tradeoff in the student's true system of meta-values W . Consequently, W is a weighted average of U_1 and U_2 , which I write $W=bU_1+(1-b)U_2$, where $0 \leq b \leq 1$. The student's uncertainty implies that he does not know the true value of b . Fortunately, a Pareto superior alternative yields higher W for all possible values of b in the interval $0 \leq b \leq 1$. Consequently, the student who chooses the Pareto superior alternative will not regret his choice later when he understands the true value of b .

Intrapersonal Pareto superiority has an important connection to cognitive dissonance. In my example cognitive dissonance theory predicts that a student who commits to accounting will develop the preferences of an accountant U_1 , and a student who commits to music will develop the preferences of a musician U_2 . If the student makes the Pareto superior choice and cognitive dissonance

⁵ A meta-preference feature with this function corresponds to a non-decreasing social welfare function written over the preferences of different individuals.

operates as predicted, the student will not regret his choice. If, however the student makes the Pareto inferior choice and cognitive dissonance operates as predicted, the student will regret his choice.

Applied to individuals, the predictions of cognitive dissonance theory are probabilities, not certainties. Over time, a person may reflect on his choices and preferences, which can undermine and even undo non-rational processes like cognitive dissonance. In the extreme case, the actual direction of change in values might reverse the prediction of cognitive dissonance theory. In the reverse case, the student who commits to accounting ends up with the musician's values and the student who commits to music ends up with the accountant's values. If the student makes the Pareto superior choice and his values change in the reverse direction from the prediction of cognitive dissonance theory, the student will not regret his choice. If, however the student makes the Pareto inferior choice and his values change in the reverse direction from the prediction of cognitive dissonance theory, the student will regret his choice.

V. Conclusion and Possible Application to Law

Economic theories of behavior especially concern choosing the best means to an end. A person treats himself as a means when he changes himself to achieve his ends. The theory of cognitive dissonance predicts how commitments, which change objective opportunities, change subjective values. By definition, a commitment to one alternative increases the objective price of the rejected alternative. By cognitive dissonance theory, commitment causes an increase in the subjective value of the chosen alternative.

Cognitive dissonance theory provides a framework to model how law changes values. Laws create obligations backed by sanctions, which raises the price of doing wrong relative to doing right. In a democracy, citizens ideally see themselves as makers of the law. When this ideal is achieved, citizens see themselves as having raised the relative price of wrongdoing. In other words, citizens ideally see their legal obligations as their commitments. When democratic commitments strengthen, the citizens like conforming to the law.

A reasonable person, who recognizes that he will probably come to like his commitments, seeks a rational basis for them. Since commitments change values, a person must choose among values in order to make commitments. A rational person may use meta-values like happiness to choose among values. In many circumstances, however, choosing by meta-values demands more self-knowledge than the decision maker possesses.

In these circumstances, the decision maker can sometimes avoid the problem if the connection between opportunities and values is strong enough. A

strong connection can cause one outcome to dominate the others when evaluated with any likely values. I call a choice that is better with respect to any likely preferences of a decision-maker “intra-personally Pareto superior” or, more simply, a “Pareto self-improvement.”⁶ According to this analysis, people will tend to change their preferences when doing so increases their satisfaction relative to their initial preferences and their final preferences.

In a well-organized democracy, citizens gain advantages from obeying the law and participating in government, relative to disobeying the law and not participating in government. Many people, consequently, commit to being good citizens. The commitment is stable because they have no intra-personally Pareto superior alternative. Segments of the population who do not gain from obeying the law and participating in government tend not to commit to being good citizens. A well-organized democracy tries to extend the advantages of obeying the law and participating in government to everyone so that everyone has a reason to commit to being a good citizen.

Cognitive psychology, which has influenced law and economic theory, is a small part of psychology. The larger part includes motivational psychology. If successful, the theory that I propose in this paper could provide a framework for incorporating motivational psychology into law and economic theory.

⁶ I analyze this concept graphically “Expressive Law and Economics,”²⁷ J. Legal Studies 585-608 I (1998) at 599; and Cooter, “Self-control and Self-improvement for the ‘Bad Man’ of Holmes” (O. W. Holmes Centenary Lecture Conference),⁷⁸ Boston University Law Review 903-930 (1998), at 923.