Abstract

Many people are suspicious of self-interest and incentives and oppose motivating humans through incentives. I analyze the meaning of incentives in the logic of choice and argue that it is inconceivable that purposeful actions are anything other than responses to incentives. Money is not always the best way to motivate people. But where money incentives are required, they are required precisely because people are motivated by things other than money.

Self-interest is consistent with altruistic motives. Agency problems, however, cannot be solved by instilling greater altruism in people because altruism, the concern for the well-being of others, does not make a person into a perfect agent who does the bidding of others.

I discuss the universal tendency of people to behave in non-rational ways. Though they are Resourceful, Evaluative Maximizers (REMMs) humans are imperfect because their brains are biologically structured so as to blind them from perceiving and correcting errors that cause psychic pain. The result is systematic, non-functional behavior. I discuss a Pain Avoidance Model (PAM) that complements REMM by capturing the non-rational component of human behavior (the crux of human self-control problems). Recognizing these self-control problems leads to an expansion of agency theory since they are a second source of agency costs in addition to those generated by conflicts of interest between people.
Self-Interest, Altruism, Incentives, and Agency Theory

Michael C. Jensen

1. Introduction

In “Incentives, Rationality, and Society” Michael Brennan (1994) attacks the use of incentives for executive compensation and, indirectly, for use in society in general. Professor Brennan argues that economics views rational behavior as self-interested and that this proposition is wrong both in a positive sense (that is, people don’t behave that way) and in a normative sense (because if they did behave in a self-interested way, the world would be a more brutish and undesirable place). Brennan is correct that people do not always behave in a rational way, but this fact provides no support for his opposition to incentives, or his opposition to self-interest.

Professor Brennan’s arguments, although popular in many quarters, are, I believe, logically and empirically incorrect. Underlying his argument is a strong aversion to incentives, an opposition that reveals itself in several places to be primarily, if not wholly, an antagonism to the use of monetary incentives. (He cites approvingly the incentives provided by awards banquets and small prizes). He is also not fond of the agency model, concluding that the model is incorrect and moreover, that “… if we go on hammering into our students the mistaken notion that rationality is identical with self-interest we shall make our agency models come true, but at the cost of producing a society that will not function.” (p. 39) These views are held by many people and therefore are worthy of additional discussion. I shall deal first with the issue of incentives and then move on to self interest, non-rational behavior, and the theory of agency.

2. The Meaning and Role of Incentives in the Logic of Choice

Professor Brennan argues disapprovingly (p. 34), that economic man “…will never perform without incentives.” Brennan appears to desire a world in which there are no incentives, and at times seems to believe that action in the absence of incentives is the natural and, therefore, desirable state of affairs. Many managers, policy makers, and religious leaders share this view.

* Edsel Bryant Ford Professor of Business Administration, Harvard Business School. This research has been supported by the Division of Research, Harvard Business School.

I am indebted to many for comments and discussions over the years that have helped me understand these issues. Although they bear no responsibility for the views expressed herein, I am grateful to Chris Argyris, George Baker, Donald Chew, Perry Fagan, Donna Feinberg, Gerald Fischbach, Jack Gabarro, Richard Hackman, Steven Hyman, William Meckling, Steve-Anna Stephens, Richard Thayler, Karen Wruck, the participants in the Mind, Brain, Behavior initiative at Harvard University, and my former colleagues at the University of Rochester.
It is inconceivable that purposeful action on the part of human beings can be viewed as anything other than responses to incentives. Indeed, the issue of incentives goes to the heart of what it means to maximize or optimize, indeed to the very core of what it means to choose. Rational individuals always choose the option that makes them better off as they see it. This is, by definition, what we mean by purposeful action—the attempt to accomplish some end. And it is the difference in (expected) well-offness between taking one action as opposed to another, that provides incentives and results in choice.

As I said, much anti-incentive reasoning seems to be based on the notion of a natural state of affairs in which there are no incentives and no conflicts of interest. For there to be no incentives, all alternative courses of action or choices must promise the same degree of utility as viewed by the individual. Such a world is virtually inconceivable (and certainly uninteresting if it did exist). It does not exist in the state of nature: not all land is equally fertile, nor are all paths equally level or unobstructed by water or other hurdles. Thus, even primitive man faced incentives to cultivate one piece of land rather than another, or to choose one path over another. Incentives exist (whether we like it or not) in all cases in which people have real choice. This is true in organizations as well as in nature.

Managers in every organization, even one with no pay-for-performance, face incentives—including incentives to do bad as well as good. To emphasize, the issue for any organization is not whether to introduce incentives to motivate its employees. There are always incentives; the issue is simply which incentives do we want to encourage and which do we want to suppress?

Optimal decisions often meet with opposition and retribution from colleagues, employees, communities, policy makers, regulators, and others with conflicting interests—thereby providing managers with incentives to forsake them. To increase the chances that managers will take the best actions possible, we must ensure that the incentives (that is, the tradeoffs) they face encourage them to move in the correct direction. This means that we must strive to ensure that the culture of the organization, and any other variables under our control, reward proper, as opposed to improper, action.

The main advantage of monetary incentives in this mosaic of organizational incentives is that general purchasing power is valued by almost everyone (because it is a claim on all resources), and it can be easily varied with performance. On the other hand, it often is difficult, and in some cases impossible, to vary other aspects of the organizational setting in a way that relates to organizational, group, or individual performance.

I hasten to add that, for a variety of reasons, monetary incentives are not the best way to motivate every action. Because monetary incentives are so strong, and because it is sometimes difficult to specify the proper performance measure, monetary incentives are often dominated by other approaches. But in the end, where money incentives are required, they are required precisely because people are motivated by things other than money.

3. Self Interest, Altruism, and Perfect Agents

History provides much evidence on the adverse consequences that follow from failing to understand the proper role of self interest, how self-interest is consistent with altruistic concerns for the
welfare of others, and how increased attention to self-interest or rational behavior would make the world a
good, not worse, place to live. Professor Brennan is correct in his observation that people do not always
behave in a rational manner. But when they do so, contrary to Brennan’s beliefs, and consistent with his
soldier and manager examples, people invariably make themselves and those around them worse off.

Professor Brennan (p. 34) defines economic man in a way often used as a caricature to discredit
economics, and I agree with him that his characterization of economic man is not descriptive of human
behavior. (Although I do not believe it characterizes the way most economists view human behavior, I shall
not belabor the point here.) People do care about failure and success, they do have emotions and care about
honor and self esteem, and they do feel shame and pride. In our paper “The Nature of Man” in this volume,
William Meckling and I discuss at length the minimal characterization of human beings capable of
capturing this complexity. We call this model REMM—for the Resourceful, Evaluative, Maximizing
Model of human behavior—and argue that these three characteristics are very powerful for describing the
rational behavior of human beings. The model is powerful in the sense that it leads to good predictions of
how human beings will actually behave. I will not repeat the arguments here.

Brennan and others err when they include altruism in the category of non-self-interested behavior.
Self-interested behavior in Brennan’s view means that people care only about themselves and not at all
about others; they have no altruistic tastes in that they are unwilling to sacrifice any of their own time or
resources for the betterment of others. This is neither logically implied by the assumption of self-interest or
empirically correct. He cites two sorts of evidence that indicates the self-interested model of human
behavior is wrong: 1) that people do evidence concern for the welfare of others, and 2) that people often do
not evidence a concern for their own welfare. I have no disagreement with the evidence.

As Meckling and I make clear in our article, there is nothing inconsistent between self-interested
and altruistic behavior. More importantly, Brennan implicitly takes the existence of altruistic behavior to
mean individuals are perfect agents for others. But the willingness to sacrifice of one’s own time, energy,
and resources for the benefit of others, by no means indicates a person is a perfect agent—a common and
dangerous fallacy. By a perfect agent I mean someone who makes decisions with no concern for his or her
own preferences, but only for those of another, including an employer or principal. As we point out in the
“Nature of Man,” there is vast evidence that people do have altruistic motives, and there is equally vast
evidence that people are not perfect agents. To find extensive evidence of altruism, we need only look to
the willingness of people to give to charity, and to help family, neighbors and even strangers.

To see equally vast evidence that people are seldom perfect agents for any interest, we need look
no farther than the conflict and abuse in families, the sexual abuse of youthful parishioners by some
Catholic priests, and the general failure of organizational control systems, (including recent abuses by the
officers and board of the United Way charity). In spite of this plentiful evidence, Brennan (p. 37) arguing in
favor of the perfect agent model says: “… we are also capable of rising above our own narrow self-interest,
and we are more likely to do so the greater is our responsibility for the welfare of others.” In the extreme, it
appears only too easy for men such as Hitler or the Romanian dictator Ceasescu to impose huge costs on
the populations with whose welfare they are nominally vested. Unfortunately, as the media coverage
illustrates in these days of race wars and regional violence, such atrocities are all too common. The abusive
cases aside, Meckling and I argue that even Mother Teresa is not a perfect agent because she would be unwilling to devote her time and energy to arbitrarily chosen ends of an employer. Her preferences evidence a strong desire to help the poor of Calcutta over all sorts of alternatives, and to that extent she is self-interested although altruistic. In this sense we are all alike in that we have our own preferences. And as I suggest in the next section of the paper, greater attention to self-interest or rational behavior would make the world a better, not worse, place to live.

4. Non-Rational Behavior

Brennan cites several examples from an article by Postrel and Rumelt (1992) of soldiers and managers behaving in ways that are clearly inconsistent with their own near-term self interest. The soldiers fail to take obvious steps (such as keeping their helmets and arms, and digging fox holes) to maintain their own protection and fighting capability, while the managers routinely mistreat customers. Such self-damaging behavior is not unique, and economists viewing the world through their lens of rationality generally fail to see that such behavior is common to virtually everyone.

While Postrel, Rumelt and Brennan are correct that people do not always act in their own self interest, this provides no support for Brennan’s call for suppression of incentives. In fact, Postrel and Rumelt correctly see the import of their stories as calling for even more precise use of incentives to address self control problems—for example, by paying attention to the timing and context of rewards and punishments to help individuals resist self-destructive impulses.

REMM, provides great predictive power and a solid basis for policy decisions in business as well as personal and public life, but it is not complete. The problem is the essentially dualistic nature of human behavior: people evidence both rational and non-rational behaviors, and these behaviors coexist in inherent contradiction. These contradictions are not aberrations from normality, but rather an integral part of normality itself. The soldier example illustrates behavior that contradicts REMM, but the contradiction goes far beyond the extreme conditions of war. Clinical psychological records as well as everyday observations of family, organizational, and social action, abound with examples of non-rational behavior.

By non-rational I do not mean random or unexplainable. I mean non-functional behavior that systematically harms the individual. But this non-rational behavior can be modeled, and I believe it is possible to integrate it with REMM, and thereby vastly increasing the predictive power of our analytical apparatus.¹

In “Economics, Organizations, and Non-Rational Behavior” (1995) I postulate a Pain Avoidance Model (PAM) to complement REMM. PAM describes non-rational human behavior that arises under

¹ To some extent, this non-optimal behavior can be incorporated into REMM by recognizing that individuals’ visions of the world and their ability to act or react depend on various factors (such as the intensity of emotions) that have the power to change, if only temporarily, their perception of “goods.” Such an expanded model could explain deviant behavior as the result of maximizing actions in situations where an individual’s perceptions of the world are systematically different, or more constrained, than normal.. Abraham Zaleznik suggested this constrained maximization view of deviant behavior. Weick’s (1983) survey of the effects of stress on the behavior of humans describes behavior that is consistent with such a view.
conditions of fear. While attempting (generally subconsciously) to avoid the pain associated with recognizing personal error, people often put themselves in a position where they incur far more pain, and end up making themselves worse off than if they had just recognized their error in the first place.

Unfortunately, these non-rational behaviors are systematic, not random. Recent brain research indicates these virtually universal human defensive responses are founded in the biological and chemical structure of the brain—responses that are intimately involved with the brain’s “fight or flight” response.\(^2\)

What is surprising about this defensive behavior is the seemingly trivial events or challenges that frighten humans, and that while they are experiencing this fright, these same individuals are commonly unaware of it. This frequently occurs, for example, when people are faced with evidence or theories that threaten to change the lens through which they view reality, including not only the world around them, but their view of themselves.

We commonly observe people refusing to learn or to change, even when the personal consequences of this refusal are negative—for example, the common tendency of managers to refuse to adapt to changes in the competitive environment in the absence of a major organizational crisis. Even scholars, whose business is the creation of new theories and knowledge, commonly react negatively (and sometimes with much personal anger) to those new theories and evidence. It is as if old ideas form ruts in our brains that prevent change.

The primary result of defensiveness is a reluctance to learn and an inability to adjust behavior to serve better our own interests. These non-rational responses are evident throughout organizations and society. Examples include the tendency of people to systematically overrate themselves in rankings of their peers, the infrequent use of pay-for-performance compensation systems, the common refusal of people to welcome feedback on their errors, and the systematic tendency for corporate managers to overpay for acquisitions.\(^3\) These non-rational responses lie, for example, at the heart of the failure of the internal control systems that has led to the waste of hundreds of billions of dollars and the failure of a number of crown jewels of corporate America over the last two decades. (See Jensen (1993).)

While grounded in the flight or fight response—a mechanism that has contributed to survival for millions of years—the human responses to emotional or psychic pain are apparently highly counter-productive in our densely populated and complex modern industrial world. This phenomenon causes people to become stuck, unchanging and unchangeable. They become wedded to the theories in their brains in ways that make them systematically worse off. And to make these challenges more difficult to deal with, the biological structure of the brain generally makes the individual blind to his or her own behavior. For this reason, learning in these situations is both very difficult and slow.

An interesting aspect of this non-rational behavior is the universal tendency for humans to hold views of themselves and their behavior—“espoused theories” in the language of Argyris (1990, 1991, 1993)—that are false. They are false because they are inconsistent with the theories on which the individuals act—what Argyris calls their “theories in use.” Discovering these violations of our own

---

\(^2\) See La Doux (1994) and the references therein, and the September 1992 *Scientific American* special issue on the brain.

\(^3\) For example, in large sample surveys, almost no one ranks themselves below the 50th percentile of their peers. For discussion of these phenomena see Baker, Jensen, and Murphy (1988), Jensen and Murphy...
principles and beliefs amounts to discovering our “irrationality,” and this is highly threatening to an individual’s self esteem. The brain’s fear response and the associated anxiety then severely limits our ability to perceive when our espoused theories are false. It is obviously difficult to change incorrect behavior under such circumstances.

The common disconnect between espoused theories and theories in use undermines the premise of Brennan’s argument for limiting students ability to learn that people generally behave in self-interested ways. He asserts, in his opening paragraph, that peoples’ espoused theories determine their behavior. “Self-image is a major determinant of behaviour. …If I think of myself as an honest person, I shall behave in an honest way.” The evidence contradicting this theory is observable around us every day.

The combination of REMM and PAM leads to a much richer positive description of behavior. Even more importantly, this dualistic model will lead to much more effective normative propositions, programs, and devices for helping people minimize the non-rational, non-functional aspects of their behavior.

Note that these tendencies for individuals to act in ways that are inconsistent with their self interests are another source of conflicts with employees, employers, partners, mates, colleagues, and so on. Therefore, they are another source of agency costs that must be addressed in any cooperative effort. And this, in turn, leads us to a structure for thinking about how to use group processes and organizational procedures to limit the counter-productive effects of this behavior.

In this expanded view of human behavior, REMM becomes more than a purely positive description of human behavior—it becomes a normative model that says this is how humans should behave. Consistent with this view, the solutions suggested by the psychological and psychiatric professions are best interpreted as helping people learn to correct their “mistakes” in order to behave in more REMM-like ways.4

The problem is not, therefore, to persuade people to avoid acting in their own self-interest as Brennan argues, but to help them learn to avoid the systematic mistakes that make themselves, and those around them, worse off. I am not advocating that people learn to behave in an unbridled self-interested way in which they care not for the welfare of others, honesty, or honor. Without honor, there can be no trust, and without trust many beneficial activities and exchanges will not occur. Honesty, honor, and trust are therefore in the self-interest of all humans.

5. Agency Theory

I now turn to agency theory, its definition, and the opposition to it from Brennan and others. In the end, of course, the usefulness of any theory depends on its ability to describe the world. No theory, including agency, is perfect, but I am surprised by the opposition that is based on a patriarchal fear that if

---

4 See Peck (1978), and Argyris (1991, 1993). Consistent with this there is some evidence that those behaving according to the REMM model live more successful lives. See Larrick, Nisbett, and Morgan (1993).
we let the masses find out how people really behave it will further corrupt them. Let me take a moment to briefly review the modern history of this topic as I have witnessed it.

In the early 1970s William Meckling and I wrote an article on the theory of agency whose purpose was to open up the black box in economics and finance called the firm.\[5\] Our point was to show that firms were highly unlikely to behave in the value maximizing way that was common to so much of the modeling and analysis in economics and finance. In opening up the box, we provided a structure within which we could begin to analyze the conflicts of interest between the involved human beings and the resulting equilibrium behavior of the organizations. Judging by the subsequent theoretical and empirical use of these concepts, the profession has found the approach a productive one.

We opened that original paper with the famous statement from Adam Smith in which he called attention to the conflicts of interest between the owners of joint stock companies and their managers and directors. Although he did not use the language of agency costs he clearly understood the problem and in a sense was the original agency theorist. Smith believed so strongly in the power of self interest and the conflicts it generates, that he was extremely pessimistic about the ability of the joint stock company to survive in any but the simplest of activities where management’s behavior could be easily monitored.

Without a monopoly…a joint stock company, it would appear from experience, cannot long carry on any branch of foreign trade. To buy in one market, in order to sell, with profit, in another, when there are many competitors in both; to watch over, not only the occasional variations in the demand, but the much greater and more frequent variations in the competition…is a species of warfare…which can scarce ever be conducted successfully, without such an unremitting exertion of vigilance and attention, as cannot long be expected from the directors of a joint stock company. (Smith, 1776, p. 460)

Smith observed the failure of dozens of stock companies in his day, and concluded that the only activities a joint stock company could carry on successfully without a state granted monopoly were those which were easily monitored—in his words in “… which all the operations are capable of being reduced to what is called a routine, or to such a uniformity of method as admits of little or no variation.” (p. 461) In his opinion this limited joint stock companies to banking, fire and marine insurance, canals, and municipal water supply. Smith did not, obviously, anticipate the evolution of governance mechanisms that have enabled the corporation to survive and even thrive in a plethora of economic activities throughout the world.

Agency theory postulates that because people are, in the end, self-interested they will have conflicts of interests over at least some issues any time they attempt to engage in cooperative endeavors. This cooperation includes not only commerce through partnerships, and corporations, but also interaction in families and other social organizations. We focused our analysis in our original article on the conflicts of interest between stockholders and managers in the public corporation, not only because of the vast extent of

---

\[5\] Jensen and Meckling (1976). The agency article was written at the same time we were working on early versions of the “Nature of Man,” and in a sense this latter paper is the foundation of our work on agency. Although the “Nature of Man” has been widely circulated over the years, it remained unpublished because it seemed to be such a straight forward summary of 200 years of work in economics. Yet the puzzle has been that the manuscript has generated considerable controversy similar to that represented in Professor Brennan’s article. The reasons, I now see, go to the heart of the conflict between science and views on the perfectibility of human beings that emanate from religious preachings as well as the secular views that come from other branches of the social sciences, and policy advocates.
the resources now controlled by such organizations, but also because those conflicts of interest are obvious and easily observed in the world around us. Similarly each of us can observe the conflicts in families, and in the social and other organizations we associate with, and the structure of our analysis applies to each of them. The conflicts are so ubiquitous and so familiar that they almost become invisible.

Meckling and I argued that because conflicts of interest cause problems and therefore losses to the parties involved, the parties themselves have strong motivation to minimize the agency costs (as we labeled them) of such cooperation. This conservation of value principle is the basic force that motivates both principal and agent, or partners, to minimize the sum of the costs of writing and enforcing (implicit and explicit) contracts through monitoring and bonding, and the residual loss incurred because it will not pay to enforce all contracts perfectly. The theory provides the structure within which to model and understand a vast array of human organizational arrangements, including incentive compensation, auditing, and bonding arrangements of all kinds. That literature is large and continues to grow.

Constrained at the time by our economists’ view of rationality, Meckling and I discussed only one source of agency costs, that which emanates from the conflicts of interest between people. There is clearly a second major source of agency costs, the costs incurred as a result of the self control problems discussed above—that is the actions that people take that harm themselves as well as those around them—what Richard Thayler years ago characterized as agency problems with one’s self.6

It is now time to take the next logical step. We have as a profession broken open the black box called the firm and are fully cognizant that, because of the inherent conflicts of interest among the parties to the firm, it doesn’t maximize in any simple sense.7 The next step is to break open the black box called the individual and similarly recognize that, because of self control problems that lead to non-functional behavior, the individual cannot be said to maximize in the simple sense we economists have assumed for 200 years. Indeed, some of the apparent contradictions evident in Adam Smith’s work, especially that between the Wealth of Nations and his Theory of Moral Sentiments, may well reflect this conflicting dualistic nature of humanity. As my discussion above indicates, self control problems are much more important than most economists have perceived. I predict they will receive vastly increased attention in the future—attention that will result in better models and policies, both public and private.

6. Some Closing Thoughts

Professor Brennan laments that the teaching of agency theory and the use of incentives will somehow “coarsen” humanity and make the world we live in a less hospitable place. Although this issue is worthy of careful research, it does not follow logically from the content of the theory, nor does his conjecture fit with my observations from 20 years of teaching this material.

---

6 See Thaler and Shefrin (1981). I am embarrassed to admit that Richard was a colleague at Rochester at the time he did this original work, and although I always found it interesting, I failed for more than a decade to see the generality and the importance of this self control issue.

7 One branch of opposition to some of this analysis came from those who believed that conflicts of interest exist, but are completely resolved by markets and competition.
When I teach agency theory I do not find students surprised at the existence of conflicts. They live
in the midst of them—and those conflicts which are not experienced personally can be shared in the daily
torrent of offerings from the world’s media. Because of the universality of conflict, the danger is that we
may take it for granted and thus fail to see it at all.

I find that students and business people are excited by the central proposition of agency theory,
and that central proposition is not that people are self-interested, or that conflicts exist. The central
proposition of agency theory is that rational self-interested people always have incentives to reduce or
control conflicts of interest so as to reduce the losses these conflicts engender. They can then share the
gains. Moreover, the theory provides a general structure to point the way to a variety of classes of solutions
to these problems.

Perhaps Professor Brennan and others object to the fact that agency theory holds out no
encouragement that human nature can be improved, and that society, organizations, and even individuals
themselves can rid themselves of these costly conflicts of interest. The reality, however, is that even if we
could instill more altruism in everyone, agency problems would not be solved. Put simply, altruism, the
concern for the well-being of others, does not turn people into perfect agents who do the bidding of others.

I am skeptical of our ability to change human nature. Thousands of years of effort by all great
religions of the world have failed to eradicate the reality of self-interest. Instead, I place my bets on the
institutional structures, contracts, and informal arrangements that we create to reduce conflict, to govern
our relations, and to increase the extent of cooperation and the benefits we reap from it. Such efforts, which
arise from self-interest, are privately rewarding and improve the quality of life in our society.
References


