

Instrumental and Economic Rationality

OVERVIEW

This chapter considers basic questions concerning the nature of instrumental rationality, and how “economic” rationality relates to our general conception of instrumental rationality. I begin in Section 1.1 by analyzing instrumental rationality: I argue that we cannot understand an *action* as rational without reference to the beliefs on which it is based. Section 1.1 thus constitutes an argument for a specific way of understanding instrumental rationality. Sections 1.2 and 1.3 examine “economic rationality,” arguing that *Homo Economicus* adds a number of additional specific features to the general idea of an instrumentally rational action.

1.1 INSTRUMENTAL RATIONALITY

What Is Instrumental Rationality?

In a sense all of economics is about rationality. Economic analysis is based on a certain conception—or, we shall see, conceptions—of rational choice. The (very) short answer to the question “what is economics?” is “the theory of rational choice, and its consequences, under constraints.” It is because this basic idea is so powerful, and

economists have developed it in such sophisticated ways, that economic approaches have come to dominate other social sciences such as political science, as well much of political philosophy. The core of the economic model—and indeed, all our thinking about rational action—is the instrumental theory of rationality.¹ This theory was developed by Thomas Hobbes and later British empiricists and became, through them, basic to the emerging science of political economy in the nineteenth century. The model has given rise to a variety of specifications. Some follow David Hume (1711–1776) and argue that all rational action is intended to satisfy desires. Others put the point in terms of satisfying preferences (which, we shall see, leads to confusions; Section 2.1), while yet others, more prevalent in political science, tend to talk of advancing one's goals or interests. To call an economic agent a "utility maximizer" is often taken to be much the same as saying that she has "purposes that her action is designed to promote."²

The most obvious interpretation of instrumental rationality might be called Rationality as Effectiveness (RE):³

RE: Alf's action is instrumentally rational if and only if ϕ -ing is an effective way for Alf to achieve his desire, goal, end, or taste, G.

However, RE is both too narrow and too broad. To see how it is too narrow, suppose that Alf is a loyal viewer of the Weather Channel, which forecasts a clear day today. Alf, though, is very cautious, so he compares this with the forecasts of the National Weather Service and with the local meteorologist, Sam the Smiling Weatherman. They concur; it is going to be a gorgeous day. On the basis of all this information, Alf goes out without an umbrella, gets soaked in a freak thunderstorm, ruins his Converse high-tops, and comes down with pneumonia. According to RE, Alf's decision not to carry an umbrella was not rational: it was anything but an effective way to achieve his goals. More generally, RE deems "not rational" any action that sets back one's goals, no matter how diligent the agent was in getting information and hedging against risks. So any risky action—such as an investment—that turns out badly runs counter to rationality: it is irrational.

That can't be right. Whether a risky action that turns out badly is irrational depends on, say, whether the agent took care to inform himself about the risk, whether he sought to minimize the risks, and so on. Rationality, including instrumental rationality, is a concept that concerns a person's cognitive processes and her choices about what to

do, and so cannot be reduced to simple effectiveness of action. This becomes even clearer when we consider the way in which RE is too broad. Suppose that Betty never bothers with weather forecasts of any sort, but consults her Ouiji board. On the day that Alf gets soaked, Betty's Ouiji board instructed her to carry an umbrella, so she kept dry. RE deems her action rational: after all, carrying the umbrella was an effective way to achieve her goal (of staying dry). But this really seems a case of dumb luck, not instrumental rationality. As Amartya Sen reminds us, the concept of rationality focuses on the relation between goals and choices between options: it asks, given your goals, what is the rational course of action *to choose*.⁴ Our concern is the quality of the choice about what to do, not simply the effectiveness of what is done. Of course to choose an option is still to *do something*: it is to buy the Honda rather than the Ford, or to go out to dinner rather than order a pizza to be delivered. But when we evaluate the rationality of what is done, our focus is not on the actual consequences of what is done (whether, in the end, the action secured the person's goals), but whether, when she chose that action over the alternatives, it was the one that, given her beliefs, *had the best prospects for satisfying her goals*. To say that Betty's decision to ϕ was instrumentally rational is not to say that in the end she achieved her goals, but at the time of choice, given what Betty believed, it was her expectation that it would satisfy them best.

The inadequacies of Rationality as Effectiveness have led some to adopt a purely subjective test of rational choice, as in Subjective Rationality (SR):

SR: Betty's action ϕ is instrumentally rational if and only if when she chose ϕ :

- (i) her choice was based on her beliefs (*B*),
- (ii) *if B* were true beliefs, ϕ -ing would have best satisfied her desire, goal, end, purpose, or taste (*G*).

Alf believed that it would not rain, so SR deems Alf's action instrumentally rational even though he was soaked despite all his precautions: *if* he was right that it was not going to rain, *then* his action would have achieved his goals. So SR absolves Alf from irrationality, and that seems correct. But SR also implies that Betty was rational when she consulted her Ouiji board; indeed, according to SR Betty would be instrumentally rational even on those days when she consults her Ouiji board, it tells her *not* to carry an umbrella, and she gets thoroughly soaked. After all, *if*

her beliefs about the accuracy of Ouiji boards were true, then doing what they say *would be* an effective way to pursue her goals. Despite this rather odd implication, subjectivist theories of rationality have appealed to many, and we can see why. If we take Betty's beliefs as given, then we can see her *reasons* for carrying an umbrella, even though they are not *good* reasons for doing so. Insofar as our only concern is to make Betty's choice to carry her umbrella intelligible,⁵ we can understand her action as displaying a sort of "instrumental rationality." Recall here the case from Sigmund Freud that I mentioned in the Introduction. Freud renders more intelligible a patient's apparently totally unintelligible sleep ceremonies by showing that, given the girl's rather odd beliefs, she chose her sleep ceremonies because, *on her view*, it helped keep her parents apart, which was her goal.

Although thoroughly subjective theories of instrumental reasoning have their attractions in social scientific explanation, almost all attempts to develop the idea of rational action have sought to justify some constraints on what constitutes reasonable beliefs that underlie an instrumental choice.⁶ An account of instrumental rationality must build in some reference to the well-groundedness of the beliefs on which the agent acts.⁷ Consider another case of Freud's, that of "Little Hans." Little Hans "refused to go out into the street because he was afraid of horses."⁸ He believed that if he went out into the street, horses would bite him. Now *if* Little Hans's beliefs were correct, then his choice not to go on to the street would seem instrumentally rational—it makes sense to avoid getting bitten by horses. But Little Hans strikes us, as he did Freud, as not being rational, since Little Hans had no good reason to believe that horses *would* bite him if he went on to the street. Thus Freud set out to uncover why Little Hans believed such an odd thing. (Freud's conjecture was that Little Hans, desiring his mother, feared retribution from his father; Little Hans transferred his fear that his father would castrate him onto horses: they would, thought Little Hans, "bite" him.) If you think there is something irrational about Little Hans's choice not to go on to the street, you do not accept a purely subjective theory of instrumental rationality.

In our rather more mundane cases, Alf seems instrumentally rational when he does not wear a raincoat, even though he fails to achieve his goals, because his beliefs about the weather (on which he acted) were well-grounded; Betty, even though she succeeds in achieving her goals, is not instrumentally rational because her beliefs about the weather are, from an epistemic point of view, terrible. And

Little Hans is irrational because the horses will not bite him; also, if he is afraid that his father will castrate him, avoiding horses will not help much. An adequate characterization of Instrumental Rationality (IR), then, must go something like this:

IR: Alf's action ϕ is instrumentally rational only if Alf chooses ϕ because he soundly believes it is the best prospect for achieving his goals, values, ends, etc.

An agent who is instrumentally rational acts in light of his goals; his decision must be based on at least minimally sound beliefs (ones that are not grossly defective from an epistemic viewpoint), and the deliberation leading to action must not be grossly defective. I shall employ the admittedly vague ideas of a "good reason" and a "sound choice" to cover these minimal epistemic requirements.

The very idea of an instrumentally rational action leads us to questions of justified belief—"epistemology." Some resist this: they would like the theory of rational action to be independent from an account of rational belief (or at least they would like the account of rational action not to depend on an independent theory of what constitutes rational belief), and so they advance "pragmatic" justifications for constraints on what constitutes a rational belief. For example, some might agree that Little Hans is instrumentally irrational because avoiding horses is a singularly ineffective way to avoid being castrated by his father. Or think about the case of the sleep ceremonies discussed in the introduction: the girl's obsession with keeping her bedstead and the quilt apart is not apt to keep her parents from having more children. *So, it might be said, holding a belief is rational only if holding that belief is an effective way to achieve your goals.* Note that this proposal simply revives Rationality as Effectiveness (RE) by applying it to belief: the act ϕ of adopting a belief B is rational if and only if it effectively achieves one's goals. Sometimes these effective beliefs will be true beliefs, and other times not. Sometimes the best way to achieve your goals may be to have false beliefs (if you want to lose weight, it is better to believe that food which is really scrumptious tastes like dirt—then you won't be tempted to eat it). This proposal, however, seems unappealing. First, it still is apt to identify the rational with dumb luck: suppose Little Hans's refusal to go out on the street so upset his father that his father left town, and left Little Hans alone with his mother. His beliefs about horses would turn out to be rational after all. Second, we seem to confront a troublesome regress. Remember, we have concluded that an instrumental reason to perform act ϕ presupposes a notion of sound (or rational, or justified) belief (B). Now, as a way of explaining

rational belief, we simply are treating "believing *B*" as another rational action. But then, to evaluate the rationality of *this* "action" ("believing *B*") we still require some notion of a rational belief—we need to know whether it is rational to believe that believing *B* would advance one's goals. So suppose we accept *B'*, the belief that believing *B* advances one's goals. Then, however, we must consider the rationality of embracing *B'*—does believing *it* effectively advance one's goals? Further appeals to the general criterion of effectively advancing one's goals will merely lead to new levels of regress of rational justification. And this because—according to Instrumental Rationality—the rationality of accepting ϕ cannot be explicated simply in terms of whether adopting it would advance one's goals (as RE suggests); we must consider whether one soundly believes adopting it would do so.

Can Goals Be Irrational?

Instrumental rationality, though, is not only about beliefs, but goals: we choose the best prospects to achieve our goals. Beliefs can be irrational—can goals be? Here most follow the spirit of David Hume:

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter.⁹

For Hume, our reason cannot tell us what to desire, so *no desire* can ever be against reason. Contemporary philosophers, however, often hedge Hume's thesis. Although Robert Nozick acknowledges that "we have no adequate theory of the substantive rationality of goals and desires, to put to rest Hume's statement, 'It is not contrary to reason to prefer the destruction of the whole world to the scratching of my finger'," he still seeks to take "a tiny step beyond Hume"¹⁰ by identifying some rational limits on desires, goals, etc. We can identify three "tiny" steps that philosophers try to take to go beyond Hume.

1. Consistency Requirements Nozick advocates a consistency requirement between desires or goals.¹¹ According to Nozick, a person shows herself to lack "rational integration" if she has some desire for *x*, yet also desires not to desire *x*. When such a "second-order" desire

(a desire directed at another desire) conflicts with a "first-order" desire (the desire for x), Nozick argues, "it is an open question" which should be given up, but it is clear that "they do not hang together well, and a rational person would prefer that this not (continue to) be the case."¹² Perhaps. Take a standard case where this story appears to make sense: a desire to get drunk. We can imagine a person with a desire to get drunk, and also with a desire to get rid of his desire to get drunk. There appears to be the sort of conflict of desires that Nozick has in mind, a conflict between a desire and a desire directed at that desire. And this, he believes, is an irrational inconsistency. Yet this is not really so clear. Take a parallel case: suppose I desire fame but I also think I would be wiser if I didn't care about popular opinion of my work, so I desire not to want fame. But I do want fame, and yet I wish I didn't. I cannot see that this alone in any way makes me irrational. To be sure, I cannot satisfy both desires, but that occurs between all sorts of desires—say, between a woman's desire to be a stay-at-home mother and to be a CEO. She wants both—she can see the attractions of both—but she cannot have both. So long as she can order these desires for purposes of choice—so long as she consistently chooses, say, to be a CEO rather than a stay-at-home mother—she is a rational chooser, even though she wants both (she always is attracted to being a stay-at-home mother, even after she achieves her dream of being a CEO). Is it any different if the conflict is not, as in this case, between two "first-order" desires (to be a CEO and to be a stay-at-home mom), but between a first-order desire (to be famous) and a "second-order" desire (to want not to desire fame)? I am, indeed, *ambivalent* about my desire for fame, but am I irrational? If, like the woman having to choose between being a CEO and staying home with her children, I can rank my two desires for choice purposes (in the end, I choose fame), I can consistently choose, though there will always be a desire that I won't satisfy. Some say this is irrational: it is like believing that the world is a sphere and that it is flat. To others it shows that we are often ambivalent creatures who, perhaps typically, are not fully satisfied with our choices, because we have a hankering after the opposite.

2. The Master Goal According to Nassau William Senior, a leading political economist of the nineteenth century, economics could assume that everyone seeks wealth because "wealth and happiness are very seldom opposed."¹³ The ultimate goal is happiness: no matter what one desires, this is a manifestation of the pursuit of happiness or pleasure. Our goals, then, have a rational structure to them: they all serve the end of the achieving the master goal. So basic was this idea to

early political economy that "utility" generally meant "pleasure." This "hedonistic" view of our ends, long discredited, has been making something of a comeback: a number of works have been published recently about how we measure happiness, and whether market economies really make people happy.¹⁴ The old criticisms, though, remain. If hedonism is intended as an account of rationality, it follows that those who aim at goals that do not serve their own happiness are *ipso facto* irrational. That seems a very restrictive notion of rational action: the action "Alf sacrificed his own happiness to help others" would be, by definition, irrational. In response to the oddness of defining such actions as intrinsically irrational, many hedonists switch ground, arguing, as did the early utilitarian Jeremy Bentham, that:

Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we *ought to do*, as well as *to determine what we shall do*. . . . They govern us in all we do, in all we say, in all we think: every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it.¹⁵

We ought to pursue pleasure *and* we must do so. If it is impossible not to pursue one's own pleasure, then we clearly can't have cases of rational action that don't aim at pleasure, since it is impossible not to aim at pleasure. This psychological claim is extremely dubious: people aim at a variety of goals, and it hard to see how they all are ways to gain pleasure. Consider helping the poor: is that a way to give you pleasure? It is said that Thomas Hobbes argued it was. Hobbes, who was famous for arguing that people only cared about themselves, was once seen giving money to a beggar. When pressed why he did so, Hobbes replied that it gave him pleasure to see the poor man happy. But was that what Hobbes really cared about? Is helping the poor, just like drinking beer, a way to gain pleasure? (Unless one gets a lot of pleasure from it, there seem to be cheaper ways to get pleasure than giving money away!) Supposing that we can and do act for other goals besides our own pleasure, then surely it is not irrational to do so. If Hobbes thought that his reply was necessary to show that he wasn't being irrational, I think it is safe to say that he was mistaken.

3. Procedural Requirements A more common way of taking a small step beyond Hume's thesis is to provide a test that distinguishes our rational from our irrational desires. Two sorts of tests are widely

supported. According to the *Autonomous Formation Test*, only desires that were formed free of undue pressure or force, free of manipulation by others, or were formed under conditions of adequate formation, are rational desires. For example, it might be argued that if a desire to smoke cigarettes has been induced by advertising and peer pressure, and so was formed in a procedurally objectionable way, it is not rational. Alternatively, according to the *Critical Reflection Test*, only desires that *could* survive a process of critical reflection are truly rational. An adherent of the Critical Reflection Test might hold that the desire to smoke is not rational because it cannot withstand critical reflection: if people reflected on smoking and the risks, it is said, they would not continue to desire to smoke.¹⁶ On this view people who smoke only maintain their desire by refusing to look at the evidence, or engaging in wishful thinking, and so on.

Such proposals are interesting, and I do not wish to dismiss them out of hand. Their weakness, though, is that while they eliminate the obvious "bad" desires, they are also effective against "good" ones. Consider the Autonomous Formation Test. The procedure by which people come to develop a desire to be moral, for example, seems to have a lot to do with a desire to please one's parents and conform to their commands—not a desire formation process that would seem to pass muster. Even the Critical Reflection Test can have counter-intuitive consequences. As many physicians know, giving patients too much information, or asking them to vividly picture the recommended procedure, may lead them to not desire things that are clearly good for them, and which they realize are good for them. Suppose a patient has a desire to have a colonoscopy, but if he really reflected on what this involved, he would no longer desire the procedure. Would this render the desire irrational? Or is that just a good reason not to think too much about some things?

1.2 "CONSUMPTION RATIONALITY" AS DIRECT SATISFACTION OF GOALS

Action That Directly Achieves Goals

It is tempting to characterize *Homo Economicus*—"economic man"—simply in terms of instrumental rationality; indeed, I shall argue later that instrumental explanations are the preferred mode for *Homo*

Economicus. However, we cannot restrict *Homo Economicus* to instrumental reasons to act, for if we do so we cannot adequately explain consumption behavior. Say, for example, that Betty is eating ice cream and she explains her behavior by proclaiming "I like ice cream." Now a hedonist (Section 1.1) might explain Betty's action by saying that it was an instrumentally effective way of achieving pleasure. If pleasure is the sole end, and if we are constructed in ways such that we always pursue pleasure and avoid pain, then indeed all action is instrumental to experiencing this mental state. However, if we reject hedonism, this interpretation of Betty's ice cream eating is strained. In general, consumption activity is noninstrumental; if ϕ is a consumption activity, it is not usefully understood as a way to achieve some goal that is distinct from ϕ . This is an important, but often ignored, point. Suppose a person has a desire for an ice cream cone; she ate one. We do not want to say that her decision to eat the ice cream cone aimed at "satisfying her desire for an ice cream cone" as if there was a distinct thing called a "desire to eat the ice cream cone" that her eating somehow assuaged—as if it were an itch that she sought to get rid of. On this "itch" theory of desires, we perform acts that satisfy desires because we actually have the aim of getting rid of the desire. But the itch theory seems wrong. Betty does not eat the ice cream cone to get rid of her "itch" to eat the ice cream cone: she eats the ice cream cone because she likes ice cream cones. The "aim" of her eating the ice cream cone is, well, to eat an ice cream cone.

This, however, raises the specter of entirely vacuous explanations of people's actions. We see Betty ϕ -ing (eating the ice cream cone) and we "explain" it by saying that "Betty's goal is to ϕ (eat the ice cream cone), so ϕ -ing (eating the ice cream cone) satisfies her goal." But now that which is doing the explaining (Betty has a goal to ϕ) and that which is to be explained (Betty ϕ -ed) look as if they are the same thing. The worry is that the only way we can really know that Betty has a goal of ϕ -ing is by observing her ϕ -ing. Our explanation would then be of the form:

Explanandum: (1) Betty ϕ s.

Explanans: (2) Betty's reason for ϕ -ing is that she has a goal to ϕ .

which looks too much like:

(3) Betty ϕ s because she is observed to ϕ .

This does not seem very promising as a way to explain her action; the *explanans* is essentially a restatement of the *explanandum*.

INSTRUMENTAL AND ECONOMIC RATIONALITY

Although it is tempting to revert back to a thoroughly instrumental account of *Homo Economicus* and say that the "aim" of our action is to "satisfy" our desire to perform that very action, we can solve the problem without resorting to the itch theory. Betty's specific act, ϕ , can be genuinely *explained* (as opposed to merely restated) if it is an instance of a general disposition to Φ , or engage in Φ -type acts—i.e., a disposition to engage in actions of a certain type. Thus, for instance, when Betty explains eating ice cream cones by saying "I like them," this counts as an explanation if Betty is generally disposed to eat ice cream cones. The explanation here is one of token and type, or specific instance and general kind. Now this sort of explanation is not altogether empty. Consider:

Explanandum: (1) Betty ϕ s.

Explanans: (2) Betty's reason for ϕ -ing (in this instance) is that she likes Φ -type.

which might be translated:

(3) Betty's reason for ϕ -ing (in this instance) is that she is observed to Φ -type.

In this case what is doing the explaining is not a restatement of the thing to be explained; the specific act is explained by showing it to be an instance of a general type. Again, one may be tempted to reduce this to an instrumental explanation: Betty ϕ s as a way to bring about her goal of Φ -typing. But this really is to get the relation of the specific act and the general type wrong: ϕ -ing is not a means to Φ -typing, it is an instance of it.

We can, then, complement the idea of instrumental rationality with what I shall call "Consumption Rationality" (CR):

CR: Betty's action ϕ is "consumptively rational" only if it is an instance of Φ -type—a general desire, value, or end of hers.

Let us, then, combine IR and CR for a preliminary account of Economic Rationality (ER):

ER: Charlie's action ϕ is economically rational only if it is
(a) instrumentally rational or (b) consumptively rational.

Note that ER provides necessary but not sufficient conditions for economic rationality. It does not, for example, determine what is economically rational for Charlie to do if he must choose, say,

between two consumptively rational acts— ϕ and ψ . We shall turn to that problem in Section 1.3.

Instrumentalism as the Preferred Explanation

Although the full explication of *Homo Economicus* must allow for both instrumental and consumptive rationality, its power and persuasiveness stems from instrumental explanations. *Homo Economicus* rests much of its claim to our allegiance on the “almost unbelievable delicacy and subtlety”¹⁷ of its analysis of the market. And this analysis shows that individuals with diverse goals engage in certain sorts of economic activity because they believe that such activity is instrumentally effective in achieving their different goals. Thus, for instance, the economic model can explain the choice of medicine as a career by citing relative costs of training, opportunity costs, etc., and showing how occupational choices are instrumentally rational in achieving an agent’s ends. For some economists the crux of the economic model is to explain how different actions in different situations are instrumentally rational to satisfying a stable set of goals.¹⁸ Choices, such as the choice of an occupation, are ways of achieving the satisfaction of stable goals in different circumstances. Such economists, then, would not explain occupational choice by appeal to the goal to be a doctor but, rather, as a way to satisfy one’s stable goals for income and prestige.

However, it is important to note that citing the brute desires or goals of people to be doctors does not abandon the notion of economic rationality: after all, people must have desires or goals if the economic model is to work. And market research, which sometimes seeks to uncover things people noninstrumentally value, is not opposed to *Homo Economicus*. But direct appeal to a desire to Φ -type as a way to explain ϕ -ing is not the preferred mode of economic explanation. We might say that it is an explanation of the last resort, for while it is not empty, neither does it explain much. The preferred explanation is to show that action ϕ is instrumentally rational for agents with a wide variety of goals, which are not directly about the merits of engaging in the general type of activity, Φ . Thus, the tendency I cited earlier to identify *Homo Economicus* with instrumental rationality is not entirely off the mark: although economically rational agents have access to other reasons besides instrumental ones, *Homo Economicus*, as a model of rational action, prefers explanations in terms of instrumental rationality.

1.3 HOMO ECONOMICUS

Thus far I have been developing a conception of instrumental rationality and what I have called "consumption" rationality. Economists typically do not concern themselves with these matters but, as I hope will become clear in the following chapters, it is a good idea to have in hand a general and nonformal understanding of economic rationality before we begin to look at ways to formalize the theory of rationality. But much more needs to be added to our informal model of economic rationality before we have described "*Homo Economicus*." "Economic man" builds on, but goes far beyond, basic Humean instrumental rationality. *Homo Economicus* pursues multiple goals, and operates on general principles that determine when one goal rather than another will be sought. Whereas the instrumentally rational agent as we have described her captures part of rational action, *Homo Economicus*, I will argue, looks more like a full-fledged model of an instrumentally rational agent. Let us consider five traits that are widely recognized as features of *Homo Economicus*.

1. More Is Better than Less Hume, it will be recalled, argued that "'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter." *Homo Economicus* definitely rejects this; here is a case where Hume's preference looks irrational. Suppose I adore Australian wine of a certain winemaker and vintage, and I am confronted with a choice: for \$25 I can have either one bottle or two bottles of that wine and vintage. And suppose I choose one bottle over two. Why might I do this? We can think of a variety of reasons. I might be drinking too much wine these days, and know that I'll drink both bottles right off, something of which my doctor might not approve. Or perhaps I am riding my bike home and really can't carry a second bottle. Note, though, that explanations like these appeal to some other goal (health) or consideration (the second bottle can't be transported). These other considerations make sense of what seems to be an irrational choice: a choice for less rather than more of something I desire. Imagine, though, that confronted with the choice between one or two bottles of my favorite wine I simply choose the one, saying I desire less of what I like to having more of it. A sign that we are confronting an irrational desire here is that it seems unintelligible why someone would want that: we need to tell more of a story to make the action sensible to us.

As Russell Hardin puts it, "[t]he simplest definition of rationality... is that one should choose more rather than less value."¹⁹ When faced with a choice between the satisfaction of goal G to degree p or the satisfaction of a G to level q , where p is greater than q , a rational agent has instrumental reason to obtain pG rather than qG . For just about any goal, if one has the option between satisfying it to a lesser and a greater extent, *Home Economicus* will choose to satisfy it to a greater degree. Notice that Hardin talks of "value" and I have put "more is better than less" in terms of goal satisfaction rather than simply "goods." One might simply say that *Homo Economicus* holds that more of any consumption *good* is better than less of it. The goal formulation, though, is both more general and, perhaps more importantly, more consistent with the assumption that more is better than less. It is more general because any good can be included as a goal: people can have a goal of consuming a good, so the goal specification can include the goods way of understanding "more is better than less." On the other hand it is something of a stretch to conceive many of the things we pursue, such as fame or the advance of science, as "goods." If the focus is on specific goods, moreover, both satiation and "lumpiness" are more of a problem. Sometimes we get as much of a specific good as we can handle. For example, we do not think eating a ten-pound steak is better than eating a two-pound steak. If we redefine our goal in terms of, say, pleasure or having a satisfying meal, then the anomaly is removed, for the bigger steak does not, presumably, make for a more satisfying meal: it does not yield a greater degree of goal fulfillment. Although we need not adopt hedonism, it probably is the case that rationality more obviously endorses "more is better than less" when goals are not understood as aiming at specific goods, but as more abstract aims, ends, and states of being. "Lumpiness" can also be a problem if we focus on specific goods; it may well be that we only have use for a good in "lumpy" increments, so that an extra amount that does not get us to the next increment is of no use. If I am building a new bicycle, two new (identical) seats are probably not better than one, since I can't use the second. But here the problem seems to be that my goal is not really better satisfied by a second seat, so the second seat does not achieve a higher level of goal satisfaction. One way to avoid these problems is to simply assume that goods are continuous (not lumpy), and that an extra increment always better satisfies our goal than a smaller increment. The good of money probably satisfies this assumption fairly well.

2. Decreasing Marginal Value We have seen that *Homo Economicus*, other things equal, will always prefer a greater over a lesser degree of goal satisfaction. However, although we always prefer the greater amount over the lesser, the increase in value diminishes from the n th to the $n + 1$ unit of a goal. This idea of diminishing marginal utility was one of the revolutionary developments in modern economics. In the words of Carl Menger, one of the economists usually credited with its discovery:

The satisfaction of every man's need for food up to the point where his life is thereby assured has the full importance of the maintenance of his life. Consumption exceeding this amount, again up to a certain point, has the importance of preserving his health (that is, his continuing well-being). Consumption extending beyond even this point has merely the importance—as observation shows—of a progressively weaker pleasure, until it finally reaches a certain limit at which the satisfaction of the need for food is so complete that every further intake of food contributes neither to the maintenance of life nor the preservation of health—nor does it give pleasure to the consumer, becoming first a matter of indifference to him, eventually a cause of pain, a danger to his health, and finally a danger to life itself.²⁰

Menger's basic insight was that the amount of extra value we place on additional increments of a good decreases as we gain more of the good. As we see from the last lines of this quote, Menger himself thought that at some point we reach satiation and beyond that we *disvalue* increments; so not only does the extra value of an increment become smaller, but at some point there is no increment at all (this then conflicts with "more is better than less," at least when applied to the good of food). If, though, we accept that "more is better than less," then we suppose that we never quite reach the point at which a person is entirely indifferent, though the extra value he achieves from the increment might be miniscule.

Somewhat surprisingly, although decreasing marginal utility is fundamental to economic analysis, its grounding is not obvious. Some treat it simply as an empirical generalization about how people do in fact typically choose rather than a basic feature of rationality itself or a deep psychological law.²¹ Hedonists can provide a more secure foundation for it, claiming that since everything we do aims at pleasure, it is a deep psychological law that the more we have of

something, the less extra pleasure we get from each additional unit: the first slice of pizza gives a lot of pleasure, the second less, the third even less. While this makes sense of decreasing marginal utility, it does so only if we accept hedonism, something we have already questioned (Section 1.1).

Another possibility is to see decreasing marginal utility as fundamental to the concept of an instrumentally rational agent with multiple goals. Thus far we have been treating instrumentally rational agents as if they had only one goal, desire, etc. Instrumentally rational agents, though, must act on a variety of goals. Now suppose one had a variety of goals but the satisfaction of each goal had no diminishing marginal effect: achieving the $n + 1000$ unit was just as important as achieving the n th. Suppose further, that goal G was the most important goal for you, and you could always do something to achieve G ; suppose G is watching 24-hour Australian Rules Football. If so, you would do nothing but watch Australian Rules Football. Agents whose goals are not characterized by decreasing marginal utility would tend to be monomaniacs: they would be apt to seek the same thing over and over and over again.²² We can, then, see decreasing marginal utility as crucial to the idea of a *rational multiple-goal pursuer* who seeks to satisfy different goals at different times. As one goal is achieved its importance tends to go down so that other goals rise to attention. To say that a goal's "utility" decreases, then, is simply to say that its relative importance decreases as it is increasingly satisfied and so at some point the agent will choose an action that secures a basic satisfaction of a less important goal over an act that yields ever more satisfaction of a more important goal. Insofar as we think an instrumentally rational agent is not a monomaniac who pursues just one thing all the time, we will have to include in our model some version of decreasing marginal utility.

Consider another example: suppose a person is deeply conservative, and she performs the same religious ritual over and over again. One might think this means that the value of performing the ritual does not decrease at the margins for her. Am I saying that she is *economically irrational*, because her choices do not reflect decreasing marginal utility of rituals? I would tend to think that her choices *do* reflect decreasing marginal utility: after all, she does not perform the ritual all day every day. Having performed her religious ritual, she then chooses to eat, work, or read. That we do something repeatedly is not a counterexample to decreasing marginal utility. I watch old movies every day, and sometimes two in a day. I put great value on

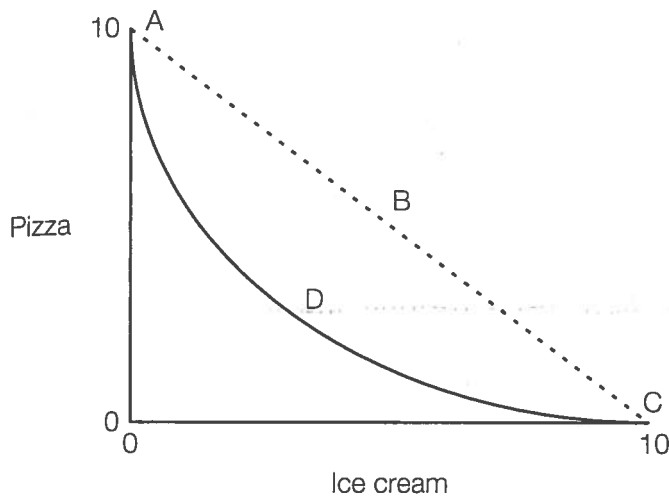


FIGURE 1-1 An Indifference Curve

them, but the point is that despite the great value I put on them, I do other things that apparently I “value less.” Why would I *ever do anything but watch old movies?* Why does our traditionalist *ever do anything other than perform her valued ritual?* Decreasing marginal utility explains why.

Much the same idea can be put in terms of *decreasing rates of substitution*. Suppose we have two goods, pizza and ice cream. And suppose Betty has \$X total to spend on some combination of pizza and ice cream. We then get Figure 1-1.

At point A, Betty spends all her money on 10 pizzas; at point C, all is spent on 10 quarts of ice cream; at B, she divides her expenditure. Given decreasing marginal utility, Betty prefers B to either A or C. This can be put in terms of decreasing rates of substitution between pizza and ice cream: when you have all pizza, you will give up quite a bit to get some ice cream; but as you get more and more ice cream, you will give up less pizza to get the same increment of additional ice cream.²³

This leads to the important idea of an *indifference curve*: Betty is *indifferent* between all bundles of ice cream and pizza on the curved line (A-D-C). Betty is, of course, not indifferent between B and A since she prefers B to A; she is, though, indifferent between A, D, and C (and this even though D is a smaller bundle of goods than A or C—because of decreasing marginal utility, D, as it were, manages to squeeze more utility out of the goods). Vilfredo Pareto (1848–1923), who is often

credited with the modern development of indifference curve analysis, described them as contours on the "hill of pleasure." Like the contours on a map, each line connects points of equal elevation, or in our case equal value; as we go up the hill, we go up in value.²⁴

3. Downward Sloping Demand Curves I have suggested that an important ground for accepting the assumption of decreasing marginal utility is that we are pursuers of multiple goals. In many ways the heart of economic thinking is that we always choose between goals (or goods). The crucial contribution of economic analysis is the idea of *opportunity costs*: the cost of getting one good thing is that we forego the opportunity to have something else that is good. This is not a lesson easily learned. When my daughter was around five years old, she would sometimes get money from a relative on a special occasion, and she would promptly go to the toy store. But it was always a tragedy for her: although she desperately wanted the toy, she desperately wanted to keep the money too. She could not bear to pay the cost of the toy, but could not bear to forego the toy either. Typically the trip ended with her breaking into tears (and me buying the toy!). The tragedy resulted because she sought an action without opportunity costs.

Given that choice always involves opportunity costs, as the cost of a choice increases it becomes less attractive to *Homo Economicus*. We can say that *Homo Economicus*'s willingness to act to satisfy goal G_1 decreases as the costs of achieving G_1 (in terms of lost opportunity to achieve G_2) increases. This is a gloss on the notion that the demand curve for any good is "downward sloping": the more a good costs, the less of it is demanded. Unlike our simple model of an instrumentally rational agent with which we began, we now see that *Homo Economicus* is a much more complicated fellow: he must be able to choose between competing actions promoting different ends through a system of trade-off rates according to which the "demand" for a goal/end decreases as its cost relative to other goals/ends increases. This is the crux of what we mean by an efficient pursuit of goals.

4. Selfishness/Wealth Maximization/ Non-Tuism *Homo Economicus* is generally understood as an egoist or, even more narrowly, a wealth maximizer.²⁵ And it is surely right that at the core of the "economic approach" to society has been the explanatory power of self-interest.²⁶ It is, of course, precisely this aspect of *Homo Economicus* that has attracted so much criticism from social scientists resisting the economic approach.²⁷ It remains, though, a matter of dispute whether

narrow motivations such as egoism or wealth maximization are really basic to the economic conception of rationality or whether they are best understood as simplifying assumptions that allow for more determinate applications of the model. The latter, I think, is more plausible. This is especially clear regarding wealth maximization: it surely is not part and parcel of being a rational agent that one maximizes wealth. However, to *assume* that people are, generally, devoted to maximizing wealth allows for determinate applications of some (not, actually, very many) economic models by allowing us to specify what the agents' goals are, and so see what actions are indeed instrumentally rational given those goals. And in many contexts it does well enough to suppose that what people are really interested in is wealth and income. However, the usefulness of this assumption is surely context-dependent. It makes sense in a few areas of economics, but most of economics does not suppose that people are wealth maximizers. Indeed, much economics does not even suppose that people are necessarily self-interested; certainly such a supposition is questionable when applied generally to political science or analysis of the law (see Chapter 6). I have tried to show here that the real crux of *Homo Economicus* can be explicated without any reference to wealth maximization or even self-interest: more is better than less, decreasing marginal value, and downward sloping demand curves all can be explained without any reference to self-interest. They simply assume that people have goals they wish to pursue, and are devoted to pursuing their own goals in the most efficient manner. Just what those goals are is another question.

Not even economics, much less all social science, must suppose that *Homo Economicus* is self-interested. According to Philip Wicksteed, "the specific character of an economic relation is not its 'egoism,' but its non-tuism."²⁸ By this Wicksteed meant that each party to an economic relation is seeking to advance her own goals, desires, and ends and not those of the other party; economics is not a study of transactions among altruists, but of non-tuists. These goals, though, may not be at all self-interested. Parents engage in economic transactions with universities: they pay large tuitions to educate their children and the universities use this money to support their own academic and social goals. Neither party need be concerned with its own self-interest. To be sure, we can save the characterization of *Homo Economicus* as inherently self-interested by describing all goal pursuit as self-interested: if it is your goal, then it is in your interest to pursue it. This, though, is to expand the meaning of "self-interest" so

that it includes anything the self is interested in. If my values lead me to give to charity, educate my children, and help my neighbors, then on this view I am being just as "self-interested" as one who keeps all her money, ignores her children, and dumps her garbage in her neighbor's yard. The idea of being "self-interested" loses its typical meaning and simply becomes just another name for non-tuism.

Even Wicksteed's "non-tuism" is merely a simplifying assumption: it allows us to calculate your "utility" as, in principle, independent of my "utility" (see Chapter 2 on the idea of "utility"). If my aim is to promote your aim, and your aim is to promote my aim, calculating what actions further our aims becomes exceedingly complicated. In such cases our utility functions are not independent, and the mathematics of modeling such interdependent functions is quite complicated. If we accept Wicksteed's "non-tuism" things are much simpler—but that is the point, it is a simplifying supposition, not something inherent to the economic understanding of rational agents.

5. Constrained Maximization The picture of *Homo Economicus* that emerges is of an instrumentally rational agent, who seeks more rather than less, who is responsive to the costs of her choices, and who does not pursue one goal again and again, but acts on a variety of goals. In this way she maximizes the achievement of her goals or ends. Economists, though, understand this maximization to occur against a background of constraints. Individuals maximize given budget constraints: in Figure 1-1, the line A-B-C is the budget constraint, showing possible combinations of goods for a fixed budget. More broadly, economic agents operate within a set of rules and institutions that constrain what they can do. That a person may maximize by attacking others is not included in most economic models, as non-coercion is usually understood as a constraint.

Need *Homo Economicus* maximize? Given her goals, must she always seek the greatest possible amount of goal satisfaction? Herbert Simon, a Nobel prize winning economist, has famously argued for an alternative to *Homo Economicus* the Maximizer—a "satisficing" agent.²⁹ Instead of seeking the best outcome—maximizing the achievement of her goals—an agent who "satisfices" seeks an outcome that is satisfactory or "good enough." Many insist that rational agents often satisfice rather than maximize. However, it is not always clear whether satisficing is really an alternative to maximizing. On one view, satisficing is simply a strategy that a maximizer might follow. People who *try* to maximize—who always seek the best—

are apt to waste a lot of time; they incur search and information costs seeing if they can do a little bit better than the “good enough” choice, and in so doing may not achieve as many of their goals as a satisficing agent, who finds a good enough option and then moves on to something else. Conceived thus, satisficing is not an alternative to maximizing, but a way to maximize (you can do best by settling for good enough and then moving on to a different goal). More radically, though, some see satisficing as an inherently rational way of acting that does not turn on a claim to ultimately maximize.³⁰ It is rational to settle for good enough even if one could, all things considered, do better. Understood in this way satisficing requires a fundamental revision of *Homo Economicus*: “more is better than less” must be abandoned, since sometimes, on this satisficing view, “good enough is as good as more.”

SUMMARY

This chapter explored the foundation of all economic analysis—the notion of a rational agent. The main aims of this chapter were:

- *To provide a defense of one conception of instrumental rationality.* Although some tend to equate “instrumental rationality” with any behavior that satisfies one’s aims, I have argued that such a view of instrumental rationality is inadequate. Instrumental rationality, I have argued, is concerned with whether, on the basis of sound beliefs, one chooses an action that has good prospects of successfully achieving one’s aims.
- *To consider whether there are good grounds for modifying Hume’s claim that reason cannot tell us what our goals should be, but can only tell us how to best achieve the goals we have.* We considered three proposals that “go beyond” Hume’s view: (1) that desires can be inconsistent, and so a rational agent only has consistent desires or aims; (2) that happiness or pleasure is the true master goal, and all other goals are only valuable insofar as they lead to happiness or pleasure; and (3) that we can identify certain tests that show which desires are rational and which are not.
- *To show that, in addition to “instrumental” rationality, an adequate account of economic rationality must explain the rationality of simple consumption choices—doing something just because you like it.*

- To sketch the ways in which the standard view of *Homo Economicus* goes beyond the simple idea of an instrumentally rational agent. I suggested that it is generally agreed that *Homo Economicus*: (1) holds that more is better than less; (2) has goals that are characterized by decreasing marginal value; and (3) has downward sloping demand curves. In Section 1.3 the important idea of an indifference curve was introduced. I also questioned (4) the widespread view that *Homo Economicus* is basically self-interested. Finally, we surveyed a debate as to whether *Homo Economicus* always seeks to maximize the satisfaction of her goals.

NOTES

1. "Instrumental rationality," says Robert Nozick, "is within the intersection of all theories of rationality (and perhaps nothing else is). In this sense it is the default theory, the theory that all can take for granted, whatever else they think. . . . The question is whether it is the whole of rationality." *The Nature of Rationality*, p. 133.
2. Geoffrey Brennan and Loren Lomasky, *Democracy and Decision: The Pure Theory of Electoral Preference*, p. 9. Note omitted.
3. For another treatment of the issues I discuss here see G.W. Mortimore, "Rational Action," pp. 96ff.
4. Amartya Sen, "Choice, Orderings and Morality," p. 55.
5. On the importance of intelligibility to economic explanations, see Alexander Rosenberg, *Philosophy of Social Science*.
6. This is true even regarding those decision theorists who describe themselves as "subjectivists." As one of the most eminent subjectivists insists, "your 'subjective' probability is not something fetched out of the sky on a whim." Richard Jeffrey, *Subjective Probability*, p. 76.
7. See further Jon Elster, "The Nature and Scope of Rational-Choice Explanation," pp. 60-72.
8. Sigmund Freud, "Inhibitions, Symptoms and Anxieties," p. 254.
9. David Hume, *A Treatise of Human Nature*, Book II, Part III, sec. 3.
10. Nozick, *The Nature of Rationality*, pp. 63, 148.
11. Nozick talks of "preferences" or goals. We shall see in Section 2.1 that "preference" is used in ambiguous ways; for now I put off employing it.
12. Nozick, *The Nature of Rationality*, p. 141.

INSTRUMENTAL AND ECONOMIC RATIONALITY

13. Nassau William Senior, *An Outline of the Science of Political Economy*, 5th edition, pp. 187–188.
14. See, for example, Bernard van Praag and Ada Ferrer-i-Carbonell, *Happiness Quantified: A Satisfaction Calculus Approach*.
15. Jeremy Bentham, *Introduction to the Principles of Morals*, p. 65. Emphasis added.
16. There are many such views. For an example, see Cass Sunstein, *Free Markets and Social Justice*, pp. 18ff.
17. George J. Stigler, *The Economist as Preacher*, p. 21.
18. See Gary Becker, *The Economic Approach to Human Behavior*.
19. Russell Hardin, *Indeterminacy and Society*, p. 16.
20. Carl Menger, *Principles of Economics*, p. 124.
21. See Daniel M. Hausman, *The Inexact and Separate Science of Economics*, p. 32. Cf. Ludwig von Mises's claim that "there is no question of any such thing as a law of increasing marginal utility" (i.e., marginal utility never increases over the range of possible quantities). *Human Action: A Treatise on Economics*, 3rd edition, p. 125.
22. See Sigmund Freud, "The Disposition to Obsessional Neurosis" pp. 134–144.
23. Decreasing marginal utility and decreasing rates of substitution are closely related but not the same idea: whereas the idea of decreasing marginal utility requires a cardinal measure of utility, decreasing rates of substitution can be expressed in purely ordinal terms. See Sections 2.2 and 2.3.
24. Those already familiar with indifference curves may find this odd: indifference curves are an ordinalist notion, but Pareto seemed to have in mind here cardinal utility qua amounts of pleasure. I explore these distinctions in Chapter 2. On Pareto's attraction to cardinal utility, see Luigino Bruni and Francesco Guala, "Vilfredo Pareto and the Epistemological Foundations of Choice Theory."
25. Brennan and Lomasky, *Democracy and Decision*, pp. 9–10.
26. See Stigler, *The Economist as Preacher*, p. 21. Cf. Gary Becker, "Altruism, Egoism and Genetic Fitness," in his *The Economic Approach to Human Behavior*, chap. 13.
27. See, for example, *Beyond Self-Interest*, edited by Jane J. Mansfield.
28. Philip H. Wicksteed, *The Common Sense of Political Economy*, vol. 1, p. 180.
29. See, for example, Herbert Simon, "Theories of Decision Making in Economics and Behavioral Science."
30. See Michael Slote, "Two Views of Satisficing," pp. 14–29.