1

Introduction: Marginal Utility Matters

1.1 THREE VIEWS OF DECISION-MAKING

In a textbook that capped a generation of work on axiomatic preference theory, David Kreps introduced Totrep or "trade-off talking rational economic person" to ease students into the mathematical models they would need to master.[1] Totrep became a celebrity, by the standards of fictional economics personalities, and served as an exemplar of the economic agent who must choose among alternative actions. Curiously the reader never learns if Totrep can pin down the marginal trade-offs that economics is famous for. Totrep's preferences must satisfy the classical axioms of rationality that hold that all pairs of alternatives can both be judged and judged consistently, but it remains open whether Totrep can determine the marginal value of one good in terms of another.

The classical axioms of rationality do not require agents to make judgments of the form "I am willing to accept x units of good 2 for a small amount of good 1 and to give up x units of good 2 to receive the same small amount of good 1." These are the judgments that underlie the first graph drawn on the blackboards of Econ 101, the smooth indifference curve that pictures an agent's marginal trade-offs of one good for another. One of the hallmarks of economics is therefore missing from the mathematical model of Totrep's preferences.

The divide between Totrep and the intro economics classroom mirrors the grand development of neoclassical economics, where two views

[1] The textbook, Kreps (1988), explains that Michael Harrison has parental rights over Totrep.

of economic decision-making have dominated the intellectual landscape and divide it into two eras of roughly equal length. In the origin story of the indifference curve in the late 19th century, marginal utility played the lead: economic agents use the pleasure delivered by increments of various goods to figure out which combinations of goods will keep them at the same level of satisfaction. For the next 75 years or so, in the marginalist period of neoclassical economics, agents were accordingly modeled by utility functions with derivatives that represent the agents' marginal utilities. The smooth indifference curve is the perennial survivor of this epoch and, to this day, the smooth indifference curve makes the most sense when it is built from the ground up by agents who weigh the increments of some primordial benefit that different options can deliver.

For the last 75 years, beginning shortly after World War II, a more spare model of rationality has ruled the seminar rooms of economic theory: it requires an agent's preferences to satisfy two axioms, completeness and transitivity, and further assumptions in more specific choice contexts. The smooth indifference curve appeared difficult to defend to the best and the brightest of postwar economic theory; it was also unnecessary for the Arrow-Debreu agenda that dominated economic theory in the initial decades of this era. The existence and optimality of competitive equilibria were the crown jewels of economic theory and, for these results, marginal utilities and marginal rates of substitution are irrelevant. Since it seemed to serve no purpose, the smooth indifference curve was abandoned by those theoretically in the know.

Much of the economics profession paid little attention to the changing of the guard. The everyday models of economics continue to rest on smooth indifference curves and differentiable utility functions, and consumer optimization is still explained to undergraduates with the story that agents equate the marginal utilities of their expenditure on different goods. The transition in economic theory also passed unnoticed in the outside world. In the public imagination, economics comes down to the maxim that "everything has its price": agents will trade away anything of value if offered enough in exchange. While this saying is something of a caricature, an agent with smooth indifference curves is remarkably malleable: if after buying positive amounts of two goods, the relative price of good 1 in terms of good 2 were to rise even slightly then the agent would happily trade away some quantity of good 1. A readiness to substitute and trade goods remains a benchmark of economic orthodoxy. But this flexibility does not follow from the axioms that describe Totrep's decisions.

There is a third position to consider. When goods do not deliver different quantities of a common homogeneous benefit, agents may be of several minds about the trade-offs they confront. Individuals can then conclude that their options are incomparable and that they are unable to come to a preference judgment: their preferences are incomplete. Incompleteness does not imply that an agent has somehow fallen prey to irrationality; the incomparability view challenges the claim that the rational pursuit of one's interests requires an agent to form preferences. Agents must still choose, of course, even when they cannot figure out which options are best. Whether facing simple or complex choices, between apples and oranges or between detailed state-contingent plans, agents may conclude that all of their conflicting attitudes must be in agreement to approve a change over the status quo or a customary decision. Or they may resort to the safest course of action, say, the plan that makes the worst-case outcome as desirable as possible. These and other choice strategies that agents turn to when they cannot form preference judgments overshadow the pleasure calculations that economists in the 19th century, eager to apply calculus, imagined to be dominant.

When agents cannot form preferences, the options they do not know how to compare cannot be grouped into conventional indifference curves: if an agent cannot compare alternatives a and b then an improvement to a need not make it superior to b. When agents resort to safe options, it may be possible for their choices to be modeled by ordered families of indifference curves, but those curves will not display the smooth trade-offs we expect of Homo Economicus. For example, the marginal value of a good might fall discontinuously as it crosses the threshold of consumption that an agent regards as safe. In both scenarios, the smooth indifference curve disappears.

Agents who lack preference judgments cannot make arbitrary choices without jeopardizing the goals such as greater material wealth that they can identify. Sticking to the status quo is the most obvious way for agents to eliminate those dangers. Seen in this light, some of the characteristic findings of behavioral economics no longer appear as inexplicable outbreaks of irrationality. Status quo bias and kindred patterns of choice lay out exactly the decisions that individuals without preferences should take to safeguard their interests. The verdict of the economics profession is that the behavioral evidence has toppled classical rationality as a positive theory of decision-making – despite its persistence in economic theory. But if the incompleteness of preferences lies behind the manifold violations of standard choice theory, then a unified explanation of economic

decision-making must go beyond the empirics of behavioral economics; a reformulation of rationality is needed.

This book will ask and answer basic questions. Which of the three views of economic decision-making is correct? Can agents always smoothly trade off disparate benefits? Why is there a discrepancy between contemporary economic theory, which has dropped the smooth indifference curve and the differentiable utility function, and the routine work of economics? Do agents obey the narrower axioms of rationality that economic theory currently backs and, if not, are they acting irrationally? How do markets perform when agents cannot make smooth trade-offs? And can government policymaking be decisive when the individuals in society are not?

It is common for economists to view the differentiable utility function as a technical convenience, not a statement of principle. In combination with the convexity of preferences, a differentiable utility allows an agent to be modeled by a system of first-order conditions, the solution of which will normally identify unique utility-maximizing demands. If instead an agent cannot form a complete set of preference judgments and thus cannot be represented by a utility function, demands are not as easy to characterize and there are multiple ways to define optimization, a morass that economists would prefer to avoid. When preference judgments are complete but utility functions fail to be differentiable, even less seems to be at stake; with some tweaks to the standard toolkit, nondifferentiable utilities can be maximized almost as easily as differentiable utilities. After going through the ritual undergraduate exercise of discovering that the demand functions for Leontief utilities appear to be well-behaved, economists mostly leave the nondifferentiable utility function behind.

The capacity of agents to trade off benefits smoothly in fact lies at the heart of conventional economics: although the Totrep axioms may omit any mention of trade-offs, the character is aptly named. But to see what trade-offs accomplish in economics, we cannot simply accept the criteria of successful model-building that the present era of economic theory has set for itself. The main results of decision and general equilibrium theory, not surprisingly, meet the tests of theoretical consistency that those traditions have laid out. In the theory of individual behavior, we instead need to examine whether rational self-interest in fact requires agents to make choices that obey the classical axioms of rationality. And we must look beyond individual optimization to the system-wide features of economic models that depend on smooth trade-offs but that general equilibrium theory has glossed over.

I will argue that the absence of smooth trade-offs leads to challenges that cannot be resolved by existing theoretical means. Before previewing this claim, let me underscore that I am not advocating a revival of the old-time religion. The smooth trade-offs and indifference curves of early neoclassical economics provided an internal theoretical coherence that the second era of economic theory has not been able to match. In terms of empirical validity, however, those assumptions and the marginal utility mechanics that lay behind them were failures and later economists have been understandably embarrassed by them. Smooth trade-offs lie at the heart of economic analysis but not of economic reality. I therefore back the third horse.

An agent that cannot pin down a marginal trade-off between goods can usually be described by a set or band of margins or supporting prices: an incremental increase in the consumption of a good will have a strictly smaller value (in terms of a comparison good) than an incremental decrease. This multiplicity of margins or valuations can be systematic, occurring not just at isolated points but at many or all consumption bundles. I will not however assume at any point in this book that agents are incapable of judging all trade-offs between goods. Agents will for example agree to part with a unit of a good when offered enough of another good in exchange. What will be missing are the marginal trade-offs and valuations that economic analysis relies on to rule over market prices and single out which government policies are optimal.

Neoclassical economics has from the outset exaggerated the importance of substitution in consumption. Economic agents do make trade-offs in consumption based in part on their preference judgments. But the magnitude of substitution may not be great enough to buffer an economy from shocks and the gains from trade that exploit differences among agents' valuations can be small. If you are seeking an explanation of the wealth of capitalist economies or of its fluctuations, substitution in consumption is not the right place to look.

The well-defined marginal rates of substitution that stem from smooth indifference curves once provided the go-to explanation of why relative prices do not move erratically through time. If instead agents are resistant to substitution and stick to particular patterns of consumption then demand will be relatively inelastic and small changes in endowments can lead relative prices to spike or plunge – a small contraction in the supply of power from the electrical grid will cause its price to jump. The neoclassical invention of smooth trade-offs assuaged these worries: the

willingness of consumers to make marginal substitutions will dampen the volatility scenarios.

Once the differentiable utility function lost its standing, economists had to find an alternative argument for the stability of relative prices over time. The answer that descended from high theory was that the endowments that generate volatility are highly unlikely to occur. This account requires an economy to begin de novo at every date with a new stock of goods and therefore does not apply to societies where goods are produced. When production is present – and enough time passes for production to affect output levels – the absence of smooth trade-offs will again lead to erratic relative prices. Production can also deliver a better explanation of what curbs price volatility: firms can transfer resources across time to tamp down the price swings that unstable individual valuations can generate.

In normative economics, determinate marginal rates of substitution play an equally pivotal role: they underlie the decisiveness of the dominant concepts of economic efficiency, both social welfare maximization and Pareto efficiency. When in contrast agents' marginal valuations are ill-defined, a wide range of policy decisions will qualify as efficient. In public goods decisions, about environmental quality for example, agents consistently declare the harm done by an incremental fouling of the environment to dwarf the value gained by an incremental clean-up. A cost-benefit test will then fail to discriminate effectively: substantial intervals of environmental quality levels will pass the test. Applied welfare economics has avoided reckoning with this paralysis by ignoring, when possible, the ample evidence that agents wield bands of marginal valuations. For the practically minded economist, the way forward has been instead to employ the smallest valuations that agents report. This footwork lets the throughput of policy recommendations flow unimpeded, but that advice will be biased against public goods.

In the welfare parables of general equilibrium theory, efficiency in an exchange economy requires there to be price lines with a common slope that *support* (are tangent to) the sets of bundles that agents prefer to their own consumption. But if the smooth indifference curve is absent and is replaced by a set of margins, the discriminatory power of this requirement collapses. Economics then loses its role of showing how to fine-tune government policies. As in the case of public goods, many and sometimes every allocation will qualify as optimal and the pursuit of efficiency will therefore lead to few nonvacuous policy recommendations. If, say, an externality appears no policy response may be called for and any

policy change that has even a minute impact on relative prices will usually fail to qualify as an efficiency improvement. When no policy can be dismissed as inefficient – even policies that every economist would judge to be distorting – economics becomes useless as a policy guide.

As with volatility, production rather than exchange paves the way forward. Increases in productivity, not the alignment of hypothetical indifference curves, drive the growth of social wealth. Technological change at the same time leads to sharp changes in the relative prices of factors and consumption goods and thus swings in the distribution of income. Economists tend to gloss over this conflict. The harm done by opening industries to productivity-enhancing competition either disappears into the black hole of distributional value judgments or is met with reassurances that injured parties can be made whole by carefully engineered compensation payments. Compensations accordingly became a centerpiece of how economic theory has dealt with the diverse repercussions of economic change. Under the best of circumstances, compensationism requires formidably detailed information about agents' preferences and trades. But with incomplete preferences, agents' decisions need not reveal their preferences; when agents are unable to judge and go for the safe option or the status quo, they may not view their selections as superior to their other alternatives. Discovering the information needed for compensation payments then becomes much harder.

The solution I propose provides an alternative design and rationale for policymaking that omits any mention of preferences. Compensations should give agents the opportunity to undertake the same trades they made previously; the policies that emerge then will not face any credible objections. When compensations based on ex ante trades are infeasible, policymakers can instead modulate the relative price changes that can undermine the fortunes of agents. A government moreover can constrain the relative prices facing households while still incentivizing efficient production via the prices that firms face; policies can thus both harness the efficiency gains of competition and avoid the price changes that inflict harm. This alternative approach can free welfare analysis from the apparent logjams where every policy option qualifies as efficient. Policymakers do need not acquiesce to the arbitrary programs and practices they inherit. A government need not stand by, for example, when technological change and international trade wreak harm on those caught on the losing side of dynamic comparative advantage; and the government's policy responses do not have to slow economic growth.

These economy-wide repercussions of missing preference judgments form Part II of this book.

Part I addresses individual decision-making. I begin by setting the 19th century dogma of marginal utility against the more parsimonious model of rationality that succeeded it in the mid-20th century, Robinson Crusoe comparing the gains of an extra minute gathering bananas or spearing fish against the completeness and transitivity axioms that model Totrep's preferences. Economic theory did not emerge unscathed from this transition. Utility and marginal utility not only allowed agents to pin down marginal trade-offs and thus find optimal decisions; they also showed that individuals can determine which of any pair of options is the better choice.

Once doubt was cast on marginal utility and pleasure-seeking, the larger principle that agents can order their options lost its justification. Without an explanatory psychology to fall back on, contemporary decision theory has remained silent on why an individual should satisfy the most basic axiom of rationality, the completeness assumption that individuals can form a preference judgment between any pair of options. In the face of this lacuna, the standing of completeness as a benchmark of rationality begins to wobble.

Agents find many decisions easy to judge. Everyone has favored clothes, foods, pastimes, and so forth. Agents will also readily come to preference judgments when choosing the best means to a known end – as when a worker opts for the highest-paying job. And difficult choices can sometimes be reduced to simpler alternatives that are easier to weigh. If say you compare two job options with disparate features – one offers higher salary and a longer commute – you may find the decision straightforward once you realize that the high-salary option will implicitly pay a trivial wage for your drive to work. But even in the simplest cases, you may not be able to pin down the marginal trade-offs essential to economics: you may reject a small return to a long commute but not be able to form sharper judgments.

Making matters worse, the comparisons that the agents of modern economics need to make are herculean. Jevons posited agents who faced small self-contained comparisons – how to allocate food on an ocean voyage for example – and he did not suggest that agents could compare disparate types of pleasure. The agents that live in current-day economic models, in contrast, must compare detailed state-contingent plans over a lifetime of consumption. But incomplete preferences do not have to stem from the complexity of decisions or a shortage of information. A

well-informed agent facing clear alternatives might not have a best choice: there may be no bedrock of true preference that lies below.

Economists have a well-rehearsed answer to claims that agents cannot form a preference between options: make agents choose and declare that their choices reveal their preferences. I will show, however, that the "revealed preferences" that emerge from such exercises will not satisfy the classical axioms of rationality, even when agents follow decision rules that never lead to dominated outcomes. An agent in short can be rational without satisfying the axioms that supposedly characterize rationality. As a body of empirical predictions, classical rationality was therefore bound to fail, though it has taken decades of documentation for that failure to be recognized.

Our era of economics has responded to the empirical defeat of rational choice theory with a shrug: "who cares what is labeled rational, what matters is behavior." This book lays out two replies, given in embryo in this chapter, first that only the rational pursuit of self-interest can explain the apparent anomalies of real-world decision-making, and second that the appraisal of social institutions depends on a valid classification of actions as rational and irrational.

There is moreover an alternative to a divorce between rationality and behavior: characterize rationality with greater precision. When individuals face static one-shot decisions, the amendments needed are relatively minor. Instead of choosing options superior to all alternatives, agents must select undominated options. Since incomplete preferences reduce the opportunities for one decision to dominate another, decision-making then becomes easier, and indeed agents may confront an embarrassment of optima. While not a wholly new phenomenon – an agent with weakly convex indifference curves can occasionally face a budget set with more than one optimum – the multiplicity that comes with incompleteness is far-reaching. Despite this difference, the mischief that incomplete preferences can cause for the static demand for goods is limited. After all, preference theory has never been able to deliver on its promise of foundations for the downward-sloping demand function; as Becker (1962) pointed out long ago, it is easier to generate well-behaved demands from irrational behavior – specifically choices uniformly distributed on budget lines – than it is from utility maximization.

The terrain is different when agents face dynamic sequences of decisions. Individuals with incomplete preferences must then take care to avoid manipulation. The simplest way for an agent to steer clear of risks is to refuse any offer to switch to an option that the agent cannot judge

relative to the pre-existing status quo, a rigidity that stands in contrast to the agents with smooth indifference curves who adapt their consumption to every relative price change. Status quo bias, the endowment effect, loss aversion – the iconic choice strategies of behavioral economics – thus emerge as validations rather than breaches of rationality. If we define agents' interests by the outcomes their decisions yield, rather than the axioms popular in economic theory, we can predict more accurately which economic behaviors will persist and which self-interest will chip away.

Incomplete preferences also resolve the puzzle of why agents so frequently fail to find a dominant option from a set of alternatives. With classical preferences, indifference is a fluke event but with incomplete preferences, an inability to judge alternatives arises systematically. In fact, once the door is open to incomplete preferences, it becomes even harder to attribute waffling to indifference: in models where agents can both be indifferent between some options and unable to form preferences for other options, indifference comes near to disappearing altogether.

The three views of decision-making adopt conflicting positions: smooth trade-offs determined by marginal utilities versus rationality axioms on preferences versus agents that cannot always come to preference judgments. The history of the contest between the first two views was written by the victors. The psychology of pleasure-seeking peddled by the early neoclassical economists appeared pointless to their mid-20th century successors and stood in the way of their scientific aspirations. Not only did the new orthodoxy hold that individual decision-making could be based on axioms of rationality rather than utility, but the smooth indifference curve appeared to be unnecessary. As I have mentioned, the features of competitive markets identified by the Arrow-Debreu model, the unifier of postwar economic theory, did not turn on marginal utilities or any of the other derivatives in the early neoclassical arsenal.[2] The labeling of neoclassical economics as marginalist was from this vantage simply a mistake. While the rear-guard defenders of utility theory put up little effective resistance, a nagging anxiety has persisted that something was lost when marginal utilities and the smooth indifference curve were dropped from the theoretical canon. One of my jobs will be to articulate this worry. We will see that the marginalist label captures part of the truth: when individuals do not substitute the satisfaction of goods at specific marginal rates, they can instead be modeled by sets of such margins.

[2] See Hahn (1961) for example.