1

Introduction

1.1 Darwin's decision problem

On a list of history's great romantics, Darwin is an unlikely candidate for inclusion. He was a man unusually devoted to the study of barnacles and, when the time came for him to decide whether to marry, he divided a sheet of paper into two vertical columns and listed the reasons for and against marriage:

MARRY	Not MARRY
Children—(if it please God)—constant companion, (friend in old age) who will feel interested in one, object to be beloved and played with—better than a dog anyhow—Home, and someone to take care of house—Charms of music and female chit-chat. These things good for one's health. Forced to visit and receive relations *but terrible loss of time*. My God, it is intolerable to think of spending one's whole life, like a neuter bee, working, working, and nothing after all.—No, no won't do.— Imagine living all one's day solitarily in smoky dirty London House.— Only picture to yourself a nice soft wife on a sofa with good fire, and books and music perhaps—compare this vision with the dingy reality of Grt Marlboro' St.	No children, (no second life) no one to care for one in old age . . . Freedom to go where one like—Choice of Society *and little of it*. Conversation of clever men at clubs.—Not forced to visit relatives, and to bend in every trifle—to have the expense and anxiety of children—perhaps quarrelling. *Loss of time*—cannot read in the evenings—fatness and idleness—anxiety and responsibility—less money for books etc—if many children forced to gain one's bread.—(But then it is very bad for one's health to work too much) Perhaps my wife won't like London; then the sentence is banishment and degradation with indolent idle fool—

At the bottom of the sheet of paper, Darwin concluded that he should indeed "Marry—Marry—Marry." Whether his closing remark was intended to be tongue-in-cheek or as a serious comment on these reflections is difficult to say: Darwin signed off on these deliberations with "Q.E.D." An interesting proof indeed (Barlow, 1987).

I suspect few people would recommend basing one's decision to marry on the outcome of such workmanlike calculations and comparisons of pros

and cons. Yet the hope that such calculations could be applied to all matters of importance – including questions underlying lifelong happiness – drove the early Utilitarians. If a "hedonic calculus" could be found, it would be a relatively simple task, they thought, to achieve the greatest good for the greatest number. One such process is in fact suggested by Darwin's decision procedure: in order to select the best outcome from a set of alternatives, simply pick any two from the set at random, determine the better of the two, discard the inferior option, and then draw a new option from the (now slightly smaller) set. Repeat this procedure until only one option remains. At the end of the process, through a simple algorithm involving only pairwise comparisons, the best of all possible options has been found.[1] Had Bentham's notion of utility values corresponded to some real, measurable quantity, they could have been used by a social planner to chart the future course of society.

Utility does not exist – at least not in the objective, measurable sense supporting the interpersonal comparisons needed for a hedonic calculus of the Benthamite kind. Hence Bentham's dream, and with it the Utilitarian project in its purest form, failed. We do not have a prudential calculus for settling important problems and, in the absence of such a calculus with methods for quantitatively comparing the real values of alternatives, Darwin's decision procedure seems to illustrate only a mere choice heuristic. Some will, no doubt, find Darwin's use of this heuristic unsettling when applied to the marriage question. Mere heuristics seem appropriate when little is at stake – such as choosing an entrée at dinner or a film to see – but for important decisions our intuitions suggest that other procedures, ones more appropriate for treating the weighty matters at hand, should be used. An appropriate procedure would give due consideration to all of the important and relevant factors of the problem at hand, such as personal values, moral principles, individual goals, and the likely causal consequences stemming from the action chosen. *These* are the factors one ought to consider when choosing, rather than merely tabulating salient features of the situation. The intuition is that serious thought and rational reflection are necessary in order to make the right decision; methods that skimp on the amount of reflection are less likely to identify the right choice. In cases where much is at stake, like marriage for example, one should think long and hard about what to do. This strikes us as common sense.

Love and marriage are difficult topics to think about from the point of view of proper decision procedures. Let us turn our attention to a subject where the

[1] One complication exists if two options may be equally good. In this case, the procedure can be extended by including a randomization device, such as a coin, to choose between two equally good options.

connection between choice and outcome is clear, unlike love and marriage, but which some people care for almost as much: chess. The game of chess has been viewed as a metaphor for life – it has three main outcomes: Win, Lose, and Draw. If one keeps score by counting the value of pieces, you can even Win Big or Lose Badly. More importantly, from our point of view, the game of chess has an optimal strategy. It can be proven that either White has a winning strategy, or Black has a winning strategy, or that each player has a strategy which guarantees at least a draw.

Chess thus seems the perfect arena in which to observe people engaged in rational, deliberative calculations that approach the ideal standard. If a perfectly rational player knew the optimal strategy to employ when playing chess, he or she would always win, or at least force a draw. Hence, one might think that the closer a player approached the ideal, deliberative standard, the better they would be at playing chess. Or, stating the connection the opposite way, one might think that the better a person is at playing chess, the closer they approach the ideal, deliberative standard. If by "ideal deliberative standard" one means a careful consideration of all the available alternatives, combined with an assessment of the respective merit of each alternative, there turns out to be very little correlation between the strength of a chess player and the breadth of their search. Although it *is* true that chess players "spend much of their time searching in the game tree for the consequences of the moves they are considering," it is also true that "the search is highly selective, attending to only a few of the multitude of possible continuations" of play (Newell and Simon, 1972, p. 750).

The author of one particular study (de Groot, 1965) attempted to measure the number of positions typically considered by grandmasters in the course of determining their next move. Surprisingly, the number of positions considered ranged between only 20 and 76, even though many more moves were possible. The narrowness of this search becomes all the more remarkable in light of the fact that, until relatively recently, people were *consistently* better chess players than computers, the very model of the "ideal, deliberative agent" which exhaustively examines all possible positions (to the extent of the computer's ability, at least). Consider the following: Gary Kasparov can evaluate roughly three chess positions a second, whereas IBM's Deep Blue can evaluate 200 000 000 a second. In the first game of the 1997 match between Kasparov and Deep Blue, each player was given three minutes per move to think. In this time, Deep Blue could examine and evalute 36 000 000 000 moves compared with Kasparov's 540. Yet Kasparov won the first match.[2]

[2] Kasparov lost the overall tournament. The results of the six-game match were as follows: Kasparov, Deep Blue, draw, draw, draw, Deep Blue.

If human chess players fall so short of the ideal deliberative standard, why do they do so well? Newell and Simon note several key differences between how people play chess and how computers play chess. First, people tend to redefine the problem they are considering.[3] Although the human player attempts to choose the best possible next move, he typically conceives of the task quite differently. Instead of simply choosing the best next move with the intention of forcing checkmate, the human player will choose the best next move which, for example, "strengthens my defensive position on the right side of the board." This ties into a second, related point: people describe and analyze chess positions using classificatory terms that are rich in their implications. The meaning attached to these classificatory terms, such as a particular board position's "vulnerability," are difficult to operationalize and translate into computational terms.

Another reason offered by Newell and Simon for the success of human chess players actually *credits* the use of heuristics. In a different experiment, Newell and Simon had players from a wide range of abilities assess a given board configuration in order to determine the future course of play. Even though the players differed greatly in their abilities and were tested separately, the set of possible moves considered tended to have considerable overlap. In particular, "the seven moves mentioned by the largest number of subjects (16 to 6) accounted for about two-thirds $\left(\frac{62}{94}\right)$ of all mentions" (Newell and Simon, 1972, p. 757). To explain this overlap, Newell and Simon speculated as follows.

> How are we to account for this high degree of consensus? First, we may look at it from a sociological standpoint. All the players, even the weakest of those studied, belong to a common chess culture. This culture is transmitted in across-the-board play, in conversation among chess players, and in writing on chess (move-by-move reports and analyses of games among grandmasters, books on chess strategy and tactics) . . . Thus, all of these players know . . . substantially all the heuristic principles that have been incorporated in existing chess programs and a great many more. They approach the position, therefore, with a common body of beliefs acquired through participation in a common culture. The beliefs are not identical, of course—else all the players would be grandmasters—but their commonality in terms of the task requirements is substantial.
>
> *(Newell and Simon, 1972, pp. 757–758)*

Players tend to focus on the same set of possible moves because they use shared heuristics to determine what the set of possible moves should be. These heuristics incorporate "a common body of beliefs acquired through participation in a common culture." Since these beliefs are based on analyses of past games, move-by-move reports, and so on, this common body of beliefs has

[3] See Newell and Simon (1972), p. 753.

considerable evidence supporting it, which justifies the adoption of that common body of beliefs. We might then refer to "common knowledge" acquired by participation in the common culture instead of just "common beliefs." Heuristics encapsulate this common knowledge in comprehensible, and readily apprehended, forms, which can then be applied in contexts different from the one in which it was originally acquired.

Life is not chess, but decision procedures used by humans in chess mimic decision procedures used by humans in life, at least in the following sense: many of the decision problems we face in real life have determinate, optimal solutions in terms of maximizing our expected payoff. If we were perfectly rational machines equipped with unlimited computational capacity, we would have little difficulty in choosing the best action to take. Since we are not these machines, we instead muddle our way through life relying on less-than-perfect calculations derived from heuristics and rules of thumb.

Given our unavoidable reliance on heuristics, any project that attempts to explain and predict individual choice in decision contexts by positing man as a perfectly rational agent appears misguided. Nevertheless, the model of man as a perfectly rational agent, the *homo economicus* so beloved by economic theorists, has been adopted by many as the standard model of the rational agent. Not everyone finds this model satisfying. In the late nineteen fifties, Herbert Simon introduced the concept of *bounded rationality* in direct opposition to the concept of perfect rationality then so commonly assumed:

> The alternative approach [to economic man] . . . is based on what I shall call the *principle of bounded rationality*:
>
> > *The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world—or even for a reasonable approximation to such objective rationality.*
>
> If the principle is correct, then the goal of classical economic theory—to predict the behavior of rational man without making an empirical investigation of his psychological properties—is unattainable. For the first consequence of the principle of bounded rationality is that the intended rationality of an actor requires him to construct a simplified model of the real situation in order to deal with it. He behaves rationally with respect to this model, and such behavior is not even approximately optimal with respect to the real world.
>
> *(Simon, 1957, pp. 198–199)*

Rejecting *homo economicus*, Simon sought to introduce a new conception of rationality, more applicable to real people, which, at the same time, improved our ability to predict the choices people make.

Simon is perhaps too pessimistic in claiming that bounded rationality is not even approximately optimal with respect to the real world. If by "approximately

optimal" one means "likely to identify the optimal choice, or a nearly optimal choice," in the case of chess, human behavior is approximately optimal. We can, and do, divide human chess players into ranked levels of ability, where a player belonging to level N can generally beat a player belonging to level $N-1$, and the probability that a player belonging to level N_+ will beat a player belonging to level N_-, where $N_+ > N_-$, rapidly converges to 1 as the distance between N_+ and N_- increases. A grandmaster will *always* beat a neophyte. Since a perfectly rational player would have the strategy that allowed him to win (or draw) regardless of his opponent, and a grandmaster can always win (or draw) when he plays people of significantly lesser ability, the heuristics used by the grandmaster are, in this sense, "approximately optimal."

Bounded rationality, in Simon's sense, means that individuals should be thought of as *satisficing* rather than *optimizing* agents. Each individual has a given *aspiration level* he wishes to attain and will take action believed to be conducive to meeting his aspiration level. Consider the problem of selling a house: the seller selects a price that she wishes to obtain, and as soon as an offer that exceeds her set price arrives, she agrees to sell the house. However, since people presumably adjust the price of a house up or down on the basis of the nature of the market, Simon allowed for the possibility that individual aspiration levels may vary in light of recently acquired information. Thus, Simon's conception of bounded rationality also includes a dynamic aspect in which people's aspiration levels vary over time.

Conceiving of people as boundedly rational agents, in Simon's sense, makes for a more descriptively accurate theory, and may even describe Darwin's deliberation reasonably well. Seeking to attain a certain level of happiness in the future, Darwin considers two courses of action and, after due deliberation, chooses marriage as more likely to make him happy. Moreover, his deliberation involves appeal to general principles best viewed as rules of thumb: he believes that marriage generally requires forced family visits and leads to a loss of time in the evenings. Even though these general principles are based on a simplified model of the real situation, Darwin entrusted his happiness to them.

The greater descriptive accuracy of bounded rationality does not mean that all vestiges of the perfectly rational agent have been removed from the theory. Gigerenzer *et al*. (1999) rightly point out that Simon omits an account of how a boundedly rational agent chooses his initial aspiration level, or how a boundedly rational agent should adjust his aspiration level in light of new evidence. In part, this is understandable since the exact procedures of adjustment will presumably be both context- and agent-dependent. One concern, though, is that many procedures for selecting an appropriate aspiration level, and for modifying the aspiration level in light of new evidence, will assume a level

of rationality that again overshoots the meagre cognitive abilities of ordinary individuals.

In order to avoid commitment to assumptions of perfect rationality, Gigerenzer *et al*. espouse a theory of "fast and frugal heuristics"[4] that downplays talk of aspiration levels and their dynamic adjustment. Rather, they argue, the heuristics people use for making decisions and taking action are efficient, easy to implement, and require minimal cognitive abilities – hence the name "fast and frugal." Such heuristics often work because they take advantage of certain structural features of the problem. As an example of such heuristics in action, consider the problem faced by a baseball outfielder who wishes to catch a fly ball in a baseball game. It seems that catching the ball requires a great deal of cognitive machinery, for the outfielder needs to infer where the ball will land given its initial trajectory and then move to that location. Determining where the ball will land, given its initial conditions, requires solving a problem in multivariable calculus. This must be done extremely quickly and accurately in order for the outfielder to have time to be at the proper location when the ball lands. As we may expect, a simpler and equally effective heuristic exists. Gigerenzer notes that, if the outfielder simply runs toward the ball so as to maintain the angle of his gaze constant, he will reach the point where the ball hits the ground at the same time as the ball arrives. This simple heuristic is extremely efficient, effective, and widely used. Gigerenzer, along with others from the Center for Adaptive Behavior and Cognition at the Max Planck Institute in Berlin, have found numerous instances in which people use other fast and frugal heuristics to reduce complex decision problems to manageable levels, many of which work surprisingly well.

In most cases, unlike the example of the outfielder and the fly ball, the heuristics provide no guarantee that the "right" answer or "best" outcome will be achieved. Each heuristic works well for a certain class of problems whose structure satisfies certain necessary conditions required for the reliable functioning of the heuristic. A heuristic recommending that an individual, when faced with a choice problem, should choose the option most recently encountered in the past will work well only when there is a correlation between the most recently encountered option and the optimality of that option. An attempt to apply such a heuristic to a new problem, for which no such correlation holds, means that, in those problem instances, the misapplied heuristic will likely perform no better than randomization, and may in fact do worse. Heuristics belong to an "adaptive toolbox" (Gigerenzer and Selten, 2001) and, just like

[4] See also Gigerenzer and Selten (2001).

tools, are guaranteed to work well only for the set of tasks they were constructed to do well. A hammer serves as a paperweight just as well as it drives nails, but the fact that the hammer performs the former function well is by accident, not design.

People are cognitively limited beings and, as such, often make choices using heuristics that *homo economicus* would scoff at. When possible, we do engage in rational deliberation, but only to the extent of which we are capable, given the limits on our abilities and the information we possess. If Gigerenzer *et al*. are correct in claiming that (1) people *do* use fast and frugal heuristics for many, if not most, of their decisions; and (2) Simon's conception of bounded rationality as satisficing requires a higher degree of rationality than does the use of fast and frugal heuristics, we should revise our reaction to Darwin's decision procedure accordingly. By approaching the marriage question as a problem of *satisficing*, we might say that Darwin did, in fact, show due respect for the solemnity of marriage. After all, he used a decision procedure that requires a higher degree of rationality than the fast and frugal heuristics he employed in other decision contexts.

1.2 Parametric and strategic choice

A careful reader, attuned to modern sensibilities, will detect an important omission from Darwin's deliberation. Whereas a great deal is made about the costs and benefits of attendant social obligations and the virtues of having children, virtually no thought is given to the possible responses by the prospective Mrs. Darwin to the marriage offer. Darwin's deliberation concerns the expected value of outcomes – Marry, Not Marry – with little regard for the connection between his choosing to get married and the actual occurrence of marriage. In short, Darwin's decision problem is one of *parametric* choice. If Darwin decides that he wants to marry, he will; if he decides that he does not want to marry, he won't. Whether the future Mrs. Darwin will accept the marriage offer, what Darwin might need to do to increase the probability of his offer's acceptance, and how the burden of these negotiations affects the overall expected benefit of being married are not subjects of consideration.

In retrospect, Darwin's framing of the marriage problem as one of parametric choice might make sense. In Victorian England few women had opportunity for meaningful careers outside of the home. For many women, ensuring a comfortable existence meant marrying well and, Darwin's love of barnacles notwithstanding, he could reasonably assume that an offer of marriage would not be turned down, provided that it was made to a woman of comparable

status. Therefore, an offer of marriage would likely lead to marriage, and so the problem really was one of parametric rather than strategic choice.

Darwin's decision problem is not ours. We live in a world where people's reactions to our choices can have a significant effect on the resulting outcome. If we recognize this, and take it seriously when deciding what to do, we soon find ourselves engaging in spiraling calculations of the form "I think that you think that I think that . . . "[5] Decision methods that work well for problems of parametric choice do not adapt well to problems of strategic choice. This point is important because problems of strategic choice tend to characterize better the choice problems faced by people in social contexts. The fact that Darwin could reasonably conceive of the marriage problem as one of parametric choice derives from peculiarities of Victorian culture more than from the nature of interdependent choice in society. Most interdependent choice problems in society have the structure of the modern marriage problem, where the outcome reflects a mutual agreement among rational persons.

The expression "a mutual agreement among rational persons" suggests the outcome of a process of rational deliberation in which all parties negotiate a settlement. Negotiating a settlement is a complex process, with many considerations by each party. Such considerations include whether one should state everything one wants from the agreement at the beginning of negotiations or refrain from stating these wants until later. The best course of action for each person depends upon what everyone else does. Interdependent decision problems of this type fall within the scope of that branch of mathematics and decision theory known as game theory.

Game theory was developed to analyze interdependent decision problems.[6] However, even though interdependent decision problems occur in many different social contexts, game theory has been, for the most part, a tool of analysis used almost solely by economists. Given the prevalence of interdependent decision problems, it is well worth asking why other disciplines have been reluctant to adopt the formal tools of game theory. I suspect part of the reason for game

[5] For example, consider the following game: members of a group of people are told to guess a number between 0 and 1, and the person whose guess is closest to the *mean* guess of the group will win $100. In the case of a tie, the money is split between those who tie. What number should a player P guess? P's guess depends upon what P thinks each other player will guess. But every other player's guess depends on what they think that P will guess. Reiterating these kinds of strategic reflections gives rise to expressions of the form "I think that you think that I think that . . . "

[6] It originated in von Neumann's seminal work on the theory of games published in 1928, and was later developed at length by von Neumann and Morgenstern in *Theory of Games and Economic Behavior* (1944). Significant resources were poured into game-theoretic research by the RAND corporation at the beginning of the Cold War. After all, what is global thermonuclear war but a game in which the only winning move is not to play?

theory's predominant confinement to economics stems from the model of the individual it employs.[7] Game theory assumes a perfectly rational agent who possesses an amazingly complete and consistent set of preferences, as well as usually perfect information about the game and the preferences of her fellow agents (although this need not always be the case). The model of man as *homo economicus* underwrites all results in game theory.

It has already been noted how real people are boundedly rational and rely on heuristics. Even with this difference between *homo economicus* and real people, one could argue that there was good reason for studying *homo economicus* as a model of real people. After all, relying on heuristics to cope with cognitive limitations effectively is a rational response. Real people might not be *homo economicus*, but they might approximate *homo economicus* in their behavior. Yet even this attempted justification for the use of *homo economicus* as a model of man faces problems, because people fail to conform to the assumptions of *homo economicus* in several important ways.

For one, people in experimental settings frequently violate the Sure Thing principle, which orthodox game theory assumes. In essence, this principle states that if a, b, and c denote possible outcomes, then one's choice between the sets $\{a, c\}$ and $\{b, c\}$ is determined solely by one's preference for a or b. On the surface, this seems plausible enough; since one gets the "sure thing" c regardless of what one chooses, it should not affect the choice.

Suppose that you are a contestant on a television game show and are told that you will play two games. The games are very simple and require no particular talent or ability. You are told that the game show's host will roll a fair, 100-sided die to determine what prize you win, and the only thing you have to do is choose what reward scheme you want before each roll of the die. That is, the choice is among reward schemes that map outcomes of the roll of the die onto personal payoffs. Suppose that, for the first game, you must choose between the following reward schemes:

p_1: receive \$100 no matter what the roll is;
p_2: receive \$0 if the host rolls a 1, \$100 if the host rolls 2–90, and \$500 if the host rolls 91–100.

For the second game, you have the following choice:

q_1: receive \$0 if the host rolls 1–89 and \$100 if the host rolls 90–100;
q_2: receive \$0 if the host rolls 1–90 and \$500 if the host rolls 91–100.

[7] It must be noted that evolutionary game theory – of considerable theoretical interest in evolutionary and population biology – adopts a very different model of the individual from that in traditional game theory, which in part explains its increasing use.