Rational Choice Explained and Defended

Herbert Gintis

March 6, 2017

1 Introduction

Choice behavior can generally be best modeled using the rational actor model, according to which individuals have a time-, state-, and social context-dependent *preference function* over outcomes, and *beliefs* concerning the probability that particular actions lead to particular outcomes. Individuals of course value outcomes besides the material goods and services depicted in economic theory. Moreover, actions may be valued for their own sake. For example, there are *character virtues*, including honesty, loyalty, and trustworthiness, that have intrinsic moral value, in addition to their effect on others or on their own reputation. Moreover, social actors generally value not only *self-regarding* payoffs such as personal income and leisure, but also *other-regarding* payoffs, such as the welfare of others, environmental integrity, fairness, reciprocity, and conformance with social norms (Gintis 2016).

The rational choice model *expresses* but does not *explain* individual preferences. Understanding the content of preferences requires rather deep forays into the psychology of goal-directed and intentional behavior (Haidt 2012), evolutionary theory (Tooby and Cosmides 1992), and problem-solving heuristics (Gigerenzer and Todd 1999). Moreover, the social actor's preference function will generally depend on his current motivational state, his previous experience and future plans, and the social situation that he faces.

The first principle of rational choice is that in any given situation, which may be time-, state-, and social-context dependent, the decision-maker, say Alice, has a *preference relation* $\succ$ over choices such that Alice prefers x to y if and only if $x \succ y$. The conditions for the existence of such a relation, developed in Section 2 below, are quite minimal, the main condition being that Alice's choices must be *transitive* in the sense that if the choice set from which Alice must choose is X with $x, y, z \in X$, then if Alice prefers x to y, and also prefers y to z, then Alice must prefer x to z as well. An additional requirement is that if Alice prefers x to y when the choice set is X, she must continue to prefer x to y in any choice set that includes both x and y. This condition can fail if the choice set itself represents a substantive social context that affects the value Alice places upon x and y. For instance, Alice may prefer fish (x) to steak (y) in a restaurant that also serves lobster (z) because the fish is likely to be very fresh in this case, whereas in a restaurant that does not serve lobster, the fish is likely to be less fresh, so Alice prefers steak (y) to fish (x). For another commonplace example, Alice may prefer a $100 sweater to a $200 sweater in a store in which the latter is the highest price sweater in the store, but might reverse her preference were the most expensive sweater in the store priced at $500. In cases such as these, a more sophisticated representation of choice sets and outcomes both satisfies the rationality assumptions and more insightfully models Alice's social choice situation.

Every argument that I have seen for rejecting the rational actor model I have found to be specious, often disingenuous and reflecting badly on the training of its author. The standard conditions for rationality, for instance, do not imply that rational Alice chooses what is in her best interest or even what gives her pleasure. There are simply *no utilitarian or instrumental implications* of these axioms. If a rational actor values giving to charity, for instance, this does not imply that he gives to charity in order to increase his happiness. A martyr is still a martyr even though the act of martyrdom may be extremely unpleasant. Nor does the analysis assume that Alice is in any sense selfish, calculating, or amoral. Finally, the rationality assumption does not suggest that Alice is "trying" to maximize utility or anything else. The maximization formulation of rational choice behavior, which we develop below, is simply an analytical convenience, akin to the least action principle in classical mechanics, or predicting the behavior of an expert billiards player by solving a set of differential equations. No one believes that light "tries to" minimize travel time, or that billiards players are brilliant differential equation solvers.

The second principle of rational choice applies when Alice's behavior involves *probabilistic* outcomes. Suppose there are a set of alternative possible *states of nature* Ω with elements $\omega_1, \ldots, \omega_n$ that can possibly materialize, and a set of outcomes X. A *lottery* is a mapping that specifies a particular outcome $x \in X$ for each state $\omega \in \Omega$. Let the set of such lotteries be $\mathcal{L}$, so any lottery $\pi \in \mathcal{L}$ gives Alice outcome $x_i = \pi(\omega_i)$ in case ω_i, where $i = 1, \ldots, n$. By our first rationality assumption, Alice has a consistent preference function over the lotteries in $\mathcal{L}$. Adding to this a few rather innocuous assumptions concerning Alice's preferences (see Section 2), it follows that Alice has a consistent preference function $u(\pi)$ over the lotteries in $\mathcal{L}$ and also Alice attaches a specific probability $p(\omega)$ to each event in Ω. This probability distribution is called Alice's *subjective prior*, or simply her *beliefs*, concerning the events in Ω. Moreover, given the preference function $u(\pi)$ and the subjective prior $p(\omega)$, Alice prefers lottery π to lottery ρ, that is $\pi \succ \rho$, precisely when the *expected utility* of π exceeds that of lottery ρ (see equation 1).

The rational actor model does not hold universally (see Section 7). There are only two substantive assumptions in the above derivation of the expected utility theorem. The first is that Alice does not suffer from *wishful thinking*. That is, the probability that Alice implicitly attaches to a particular outcome by her preference function over lotteries does not depend on how much she stands to gain or lose should that outcome occur. This assumption is certainly not always justified. For instance, believing that she might win the state lottery may give Alice more pleasure while waiting for it to happen than the cost of buying the lottery ticket. Moreover, there may be situations in which Alice will *underinvest* in a desirable outcome unless she inflates the probability that the investment will pay off (Benabou and Tirole 2002). In addition, Alice may be substantively irrational, having excessive confidence that the world conforms to her ideological preconceptions.

The second substantive assumption is that the state of nature that materializes is not affected by Alice's choice of a lottery. When this fails the subjective prior must be interpreted as a *conditional probability*, in terms of which the expected utility theorem remains valid (Stalnaker 1968). This form of the expected utility theorem is developed in Section 2.

Of course, an individual may be rational in this decision-theoretic sense, having consistent preferences and not engaging in wishful thinking, and still fail to conform to higher canons of rationality. Alice may, for instance, make foolish choices that thwart her larger objectives and threaten her well-being. She may be poorly equipped to solve challenging optimization problems. Moreover, being rational in the decision-theoretic sense does not imply that Alice's beliefs are in any way reasonable, or that she evaluates new evidence in an insightful manner.

The standard axioms underlying the rational actor model are developed in von Neumann and Morgenstern (1944) and Savage (1954). The plausibility and generality of these axioms are discussed in Section 2, where we replace Savage's assumption that beliefs are purely personal "subjective probabilities" with the notion that the individual is embedded in a *network of social actors* over which information and experience concerning the relationship between actions and outcomes is spread. The rational actor thus draws on a network of beliefs and experiences distributed among the social actors to which he is informationally and socially connected. By the sociological principle of *homophily*, social actors are likely to structure their network of personal associates according to principles of social similarity, and to alter personal tastes in the direction of increasing compatibility with networked associates (McPherson et al. 2001, Durrett and Levin 2005, Fischer et al. 2013).

It is important to understand that the rational actor model says *nothing* about how individuals form their subjective priors, or in other words, their *beliefs*. This model does say that whatever their beliefs, new evidence should induce them to transform their beliefs to be more in line with this evidence. Clearly there are many beliefs that are so strong that such updating does not occur. If one believes that something is true with probability one, then no evidence can lead to the Bayesian updating of that belief, although it could lead the individual to revise his whole belief system (Stalnaker 1996). More commonly, strong believers simply discount the uncomfortable evidence. This is no problem for the rational actor model, which simply depicts behavior rather than showing that rational choice leads to objective truth.

2 The Axioms of Rational Choice

The word *rational* has many meanings in different fields. Critics of the rational actor model almost invariably attach meanings to the term that lie quite outside the bounds of rationality as used in decision theory, and incorrectly reject the theory by referring to these extraneous meanings. We here present a set of axioms, inspired by Savage (1954), that are sufficient to derive the major tools of rational decision theory, the so-called *expected utility theorem*.[1]

A *preference function* $\succeq$ on a choice set Y is a binary relation, where $\{x \succeq y|Y\}$ is interpreted as the decision-maker weakly preferring x to y when the choice set is Y and $x, y \in Y$. By "weakly" we mean that the decision-maker may be indifferent between the two. We assume this binary relation has the following three properties, which must hold for any choice set Y, for all $x, y, z \in Y$, and for any set $Z \subset Y$:

1. *Completeness*: $\{x \succeq y|Y\}$ or $\{y \succeq x|Y\}$;
2. *Transitivity*: $\{x \succeq y|Y\}$ and $\{y \succeq z|Y\}$ imply $\{x \succeq z|Y\}$;
3. *Independence of irrelevant alternatives*: For $x, y \in Z$, $\{x \succeq y|Z\}$ if and only if $\{x \succeq y|Y\}$.

Because of the third property, we need not specify the choice set and can simply write $x \succeq y$. We also make the rationality assumption that the actor chooses his most preferred alternative. Formally, this means that given any choice set A, the individual chooses an element $x \in A$ such that for all $y \in A$, $x \succeq y$. When $x \succeq y$,

[1] I regret using the term "utility" which suggests incorrectly that the theorem is related to philosophical utilitarianism or that it presupposes that all human motivation is aimed at maximizing pleasure or happiness. The weight of tradition bids us retain the venerable name of the theorem, despite its connotational baggage.

we say "x is weakly preferred to y" because the actor can actually be indifferent between x and y.

One can imagine cases where completeness might fail. For instance an individual may find all alternatives so distasteful that he prefers to choose none of them. However, if "prefer not to choose" is an option, it can be added to the choice set with an appropriate outcome. For instance, in the movie *Sophie's Choice*, a woman is asked to choose one of her two children to save from Nazi extermination. The cost of the option "prefer not to choose" in this case was having both children exterminated.

Note that the decision-maker may have absolutely no grounds to choose x over y, given the information he possesses. In this case we have *both* $x \succeq y$ and $y \succeq x$. In this case we say that the individual is *indifferent* between x and y and we write $x \sim y$. This notion of indifference leads to a well-known philosophical problem. If preferences are transitive, then it is easy to see that indifference is also transitive. However it is easy to see that because humans have positive sensory thresholds, indifference cannot be transitive over many iterations. For instance, I may prefer more milk to less in my tea up to a certain point, but I am indifferent to amounts of milk that differ by one molecule. Yet starting with one teaspoon of milk and adding one molecule of milk at a time, eventually I will experience an amount of milk that I prefer to one teaspoon.

The transitivity axiom is implicit in the very notion of rational choice. Nevertheless, it is often asserted that intransitive choice behavior is observed (Grether and Plott 1979, Ariely 2010). In fact, most such observations satisfy transitivity when the state dependence (see Gintis 2007 and Section 3 below), time dependence (Ahlbrecht and Weber 1995, Ok and Masatlioglu 2003), and/or social context dependence (Brewer and Kramer 1986, Andreoni 1995, Cookson 2000, Carpenter et al. 2005) of preferences are taken into account.

Independence of Irrelevant Alternatives fails when the relative value of two alternatives depends on other elements of the choice set Y, but as suggested above, the axiom can usually be restored by suitably redefining the choice situation (Gintis 2009).

The most general situation in which the Independence of Irrelevant Alternatives fails is when the choice set supplies independent information concerning the *social frame* in which the decision-maker is embedded. This aspect of choice is analyzed in Section 5, where we deal with the fact that preferences are generally state-dependent; when the individual's social or personal situation changes, his preferences will change as well. Unless this factor is taken into account, rational choices may superficially appear inconsistent.

When the preference relation $\succeq$ is complete, transitive, and independent from irrelevant alternatives, we term it *consistent*. It should be clear from the above that preference consistency is an extremely weak condition that is violated only when the decision-maker is quite lacking in reasonable principles of choice.

If $\succeq$ is a consistent preference relation, then there will always exist a utility function such that individuals behave as if maximizing their utility functions over the sets Y from which they are constrained to choose. Formally, we say that a utility function $u : Y \to \mathbf{R}$ *represents* a binary relation $\succeq$ if, for all $x, y \in Y$, $u(x) \geq u(y)$ if and only if $x \succeq y$. We have the following theorem, whose simple proof we leave to the reader.

THEOREM 1. *A binary relation $\succeq$ on the finite set Y can be represented by a utility function $u : Y \to \mathbf{R}$ if and only if $\succeq$ is consistent.*

As we have stressed before, the term "utility" here is meant to have no utilitarian connotations.

3 Choice Under Uncertainty

We now assume that an action determines a *statistical distribution* of possible outcomes rather than a single particular outcome. Let X be a finite set of outcomes and let $\mathcal{A}$ be a finite set of actions. We write the set of pairs (x, a) where x is an outcome and a is an action as $X \times \mathcal{A}$. Let $\succeq$ be a consistent preference relation on $X \times \mathcal{A}$; i.e., the actor values not only the outcome, but the action itself. By theorem 1 we can associate $\succeq$ with a utility function $u : X \times \mathcal{A} \to \mathbf{R}$.

Let Ω be a finite set of *states of nature*. For instance, Ω could consist of the days of the week, so a particular state $\omega \in \Omega$ can take on the values Monday through Sunday, or Ω could be the set of permutations (about 8×10^{67} in number) of the 52 cards in a deck of cards, so each $\omega \in \Omega$ would be a particular shuffle of the deck. We call any $A \subseteq \Omega$ an *event*. For instance, if Ω is the days of the week, the event "weekend" would equal the set {Saturday, Sunday}, and if Ω is the set of card deck permutations, the event "the top card is a queen" would be the set of permutations (about 6×10^{66} in number) in which the top card is a queen.

Following Savage (1954) we show that if the individual has a preference relation over lotteries (functions that associate states of nature $\omega \in \Omega$ with outcomes $x \in X$) that has some plausible properties, then not only can the individual's preferences be represented by a preference function, but also we can infer the probabilities the individual implicitly places on various events (his so-called *subjective priors*), and the expected utility principle holds for these probabilities.

Let $\mathcal{L}$ be a set of lotteries, where a *lottery* is now a function $\pi : \Omega \to X$ that associates with each state of nature $\omega \in \Omega$ an outcome $\pi(\omega) \in X$. We suppose that the individual chooses among lotteries without knowing the state of nature, after which the state $\omega \in \Omega$ that he obtains is revealed, so that if the individual chooses action $a \in A$ that entails lottery $\pi \in \mathcal{L}$, his outcome is $\pi(\omega)$, which has payoff $u(\pi(\omega), a)$.

Now suppose the individual has a preference relation $\succ$ over $\mathcal{L} \times \mathcal{A}$. That is, the individual values not only the lottery, but the action that leads to a particular lottery. We seek a set of plausible properties of $\succ$ that together allow us to deduce (a) a utility function $u : \mathcal{L} \times \mathcal{A} \to \mathbf{R}$ corresponding to the preference relation $\succ$ over $X \times \mathcal{A}$; and (b) there is a probability distribution $p : \Omega \to \mathbf{R}$ such that the expected utility principle holds with respect to the preference relation $\succ$ over $\mathcal{L}$ and the utility function $u(\cdot, \cdot)$; i.e., if we define

$$\mathbf{E}_\pi[u|a; p] = \sum_{\omega \in \Omega} p(\omega) u(\pi(\omega), a), \tag{1}$$

then for any $\pi, \rho \in \mathcal{L}$ and any $a, b \in A$,

$$(\pi, a) \succ (\rho, b) \Longleftrightarrow \mathbf{E}_\pi[u|a; p] > \mathbf{E}_\rho[u|b; p]. \tag{2}$$

A set of axioms that ensure (2), which is called the *expected utility principle*, is formally presented in Gintis (2009). Here I present these axioms more descriptively and omit a few uninteresting mathematical details. The first condition is the rather trivial assumption that

A1. If π and ρ are two lotteries, then whether $(\pi, a) \succ (\rho, b)$ is true or false depends only on states of nature where π and ρ have different outcomes.

This axiom allows us to define a *conditional preference* $\pi \succ_A \rho$, where $A \subseteq \Omega$, which we interpret as "π is strictly preferred to ρ, conditional on event A." We define the conditional preference by revising the lotteries so that they have the same outcomes when $\omega \notin A$. Because of axiom **A1**, it does not matter what we assign to the lottery outcomes when $\omega \notin A$. This procedure also allows us to define $\succeq_A$ and $\sim_A$ in a similar manner. We say $\pi \succeq_A \rho$ if it is false that $\rho \succ_A \pi$, and we say $\pi \sim_A \rho$ if $\pi \succeq_A \rho$ and $\rho \succeq_A \pi$.

The second condition is equally trivial, and says that a lottery that gives an outcome with probability one is valued the same as the outcome:

A2. If π pays x given event A and action a, and ρ pays y given event A and action b, and if $(x, a) \succ (y, b)$, then $\pi \succ_A \rho$, and conversely.

The third condition asserts that the decision-maker's subjective prior concerning likelihood that an event A occurs is *independent* from the payoff one receives when A occurs. More precisely, let A and B be two events, let (x, a) and (y, a) be two available choices, and suppose $(x, a) \succ (y, a)$. Let π be a lottery that pays x when action a is taken and $\omega \in A$, and pays some z when $\omega \notin A$. Let ρ be a lottery that pays y when action a is taken and $\omega \in B$, and pays z when $\omega \notin B$. We say event A is *more probable than* event B, given x, y, and a if $\pi \succ \rho$. Clearly this criterion does not depend on the choice of z, by **A1.** We assume a rather strong condition:

A3. If A is more probable than B for some x, y, and a, then A is more probable than B for any other choice of x, y, and a.

This axiom, which might be termed the *no wishful thinking condition*, is often violated when individuals assume that states of nature tend to conform to their ideological preconceptions, and where they reject new information to the contrary rather than update their subjective priors (Risen 2015). Such individuals may have consistent preferences, which is sufficient to model their behavior, but their wishful thinking often entails pathological behavior. For instance, a healthy individual may understand that a certain unapproved medical treatment is a scam, but change his mind when he acquires a disease that has no conventional treatment. Similarly, an individual may attribute his child's autism to an immunization injection and continue to believe this in the face of extensive evidence concerning the safety of the treatment.

The fourth condition is another trivial assumption:

A4. Suppose the decision-maker prefers outcome x to any outcome that results from lottery ρ. Then the decision-maker prefers a lottery π that pays x with probability one to ρ.

We then have the following expected utility theorem:

THEOREM 2. *Suppose A1–A4 hold. Then there is a probability function p on the state space Ω and a utility function $u : X \rightarrow \mathbf{R}$ such that for any $\pi, \rho \in \mathcal{L}$ and any $a, b \in \mathcal{A}$, $(\pi, a) \succ (\rho, b)$ if and only if $\mathbf{E}_\pi[u|a; p] > \mathbf{E}_\rho[u|b; p]$.*

We call the probability p the individual's *subjective prior* and say that A1–A4 imply *Bayesian rationality*, because they together imply Bayesian probability updating. Because only A3 is problematic, it is plausible to accepted Bayesian rationality except in cases where some form of wishful thinking occurs, although there are other, rather exceptional, circumstances in which the expected utility theorem fails (Machina 1987, Starmer 2000).

4 Bayesian Updating with Radical Uncertainty

The only problematic axiom among those needed to demonstrate the expected utility principle is the "wishful thinking" axiom A3. While there are doubtless many cases in which at least a substantial minority of social actors engage in wishful thinking, there is considerable evidence that Bayesian updating is a key neural mechanism permitting humans to acquire complex understandings of the world given severely underdetermining data (Steyvers et al. 2006).

For instance, the spectrum of light waves received in the eye depends both on the color spectrum of the object being observed and the way the object is illuminated. Therefore inferring the object's color is severely underdetermined, yet we manage to consider most objects to have constant color even as the background illumination changes. Brainard and Freeman (1997) show that a Bayesian model solves this problem fairly well, given reasonable subjective priors as to the object's color and the effects of the illuminating spectra on the object's surface.

Several students of developmental learning have stressed that children's learning is similar to scientific hypothesis testing (Carey 1985, Gopnik and Meltzoff 1997), but without offering specific suggestions as to the calculation mechanisms involved. Recent studies suggest that these mechanisms include causal Bayesian networks (Glymour 2001, Gopnik and Schultz 2007, Gopnik and Tenenbaum 2007). One schema, known as *constraint-based learning*, uses observed patterns of independence and dependence among a set of observational variables experienced under different conditions to work backward in determining the set of causal structures compatible with the set of observations (Pearl 2000, Spirtes et al. 2001). Eight-month-old babies can calculate elementary conditional independence relations well enough to make accurate predictions (Sobel and Kirkham 2007). Two-year-olds can combine conditional independence and hands-on information to isolate causes of an effect, and four-year-olds can design purposive interventions to gain relevant information (Glymour et al. 2001, Schultz and Gopnik 2004). "By age four," observe Gopnik and Tenenbaum (2007), "children appear able to combine prior knowledge about hypotheses and new evidence in a Bayesian fashion" (p. 284). Moreover, neuroscientists have begun studying how Bayesian updating is implemented in neural circuitry (Knill and Pouget 2004).

For instance, suppose an individual wishes to evaluate a hypothesis h about the natural world given observed data x and under the constraints of a background repertoire T. The value of h may be measured by the Bayesian formula

$$\mathrm{P}_T(h|x) = \frac{\mathrm{P}_T(x|h)\mathrm{P}_T(h)}{\sum_{h'\in T}\mathrm{P}_T(x|h')\mathrm{P}_T(h')}. \tag{3}$$

Here, $\mathrm{P}_T(x|h)$ is the likelihood of the observed data x, given h and the background theory T, and $\mathrm{P}_T(h)$ gives the likelihood of h in the agent's repertoire T. The constitution of T is an area of active research. In language acquisition, it will include predispositions to recognize certain forms as grammatical and not others. In other cases, T might include physical, biological, or even theological heuristics and beliefs.

5 State-Dependent Preferences

Preferences are obviously state-dependent. For instance, Bob's preference for aspirin may depend on whether or not he has a headache. Similarly, Bob may prefer salad to steak, but having eaten the salad, he may then prefer steak to salad. These state-dependent aspects of preferences render the empirical estimation of preferences somewhat delicate, but they present no theoretical or conceptual problems.

We often observe that an individual makes a variety of distinct choices under what appear to be identical circumstances. For instance, an individual may vary his breakfast choice among several alternatives each morning without any apparent pattern to his choices. Is this a violation of rational behavior? Indeed, it is not.

Following Luce and Suppes (1965) and McFadden (1973), I represent this situation by assuming the individual has a utility function over bundles $x \in X$ of the form

$$u(x) = v(x) + \epsilon(x) \tag{4}$$

where $v(x)$ is a stable underlying utility function and $\epsilon(x)$ is a random error term representing the individual's current idiosyncratic taste for bundle x. This utility function induces a probability distribution π on X such that the probability that the individual chooses x is given by

$$p_x = \pi\{x \in X | \forall y \in X, v(x) + \epsilon(x) > v(y) + \epsilon(y)\}.$$

We assume $\sum_x p_x = 1$, so the probability that the individual is indifferent between choosing two bundles is zero. Now let $B = \{x \in X | p_x > 0\}$, so B is the set of bundles chosen with positive probability, and suppose B has at least three elements. We can express the Independence of Irrelevant Alternatives in this context by the assumption (Luce 2005) that for all $x, y \in B$,

$$\frac{p_{yx}}{p_{xy}} = \frac{\mathrm{P}[y|\{x,y\}]}{\mathrm{P}[x|\{x,y\}]} = \frac{p_y}{p_x}.$$

This means that the relative probability of choosing x vs. y does not depend on whatever other bundles are in the choice set. Note that $p_{xy} \neq 0$ for $x, y \in B$. We then have

$$p_y = \frac{p_{yz}}{p_{zy}} p_z \tag{5}$$

$$p_x = \frac{p_{xz}}{p_{zx}} p_z, \tag{6}$$

where $x, y, z \in B$ are distinct, by the Independence of Irrelevant Alternatives. Dividing the first equation by the second in (5), and noting that $p_y/p_x = p_{yx}/p_{xy}$, we have

$$\frac{p_{yx}}{p_{xy}} = \frac{p_{yz}/p_{zy}}{p_{xz}/p_{zx}}. \tag{7}$$

We can write

$$1 = \sum_{y \in B} p_y = \sum_{y \in B} \frac{p_{yx}}{p_{xy}} p_x,$$

so

$$p_x = \frac{1}{\sum_{y \in B} p_{yx}/p_{xy}} = \frac{p_{xz}/p_{zx}}{\sum_{y \in B} p_{yx}/p_{zy}}, \tag{8}$$

where the second equality comes from (7).

Let us write

$$w(x, z) = \beta \ln \frac{p_{xz}}{p_{zx}},$$

so (8) becomes

$$p_{x,B} = \frac{e^{\beta w(x,z)}}{\sum_{y \in B} e^{\beta w(y,z)}}. \tag{9}$$

But by the Independence of Irrelevant Alternatives, this expression must be independent of our choice of z, so if we write $w(x) = \ln p_{xz}$ for an arbitrary $z \in B$, we have

$$p_x = \frac{e^{\beta w(x)}}{\sum_{y \in B} e^{\beta w(y)}}. \tag{10}$$

Note that there is one free variable, β, in (10). This represents the degree to which the individual is relatively indifferent among the alternatives. As $\beta \to \infty$, the individual chooses his most preferred alternative with increasing probability, and with probability one in the limit. As $\beta \to 0$, the individual becomes more indifferent to the alternative choices.

This model helps explain the compatibility of the *preference reversal* phenomenon (Lichtenstein and Slovic 1971, Grether and Plott 1979, Tversky et al. 1990, Kirby and Herrnstein 1995, Berg et al. 2005) with the rationality postulate. As explained in Gintis (2007), in the cases discussed in the experimental literature, the experimenters offer only alternative lotteries with expected values that are very close to being equal to one another. Thus decision-makers are virtually indifferent among the choices based on the expected return criterion, so even a small influence of the social frame in which the experimenters embed the choice situation on the subjects' preference state may strongly affect their choices. For experimental support for this interpretation, see Sopher and Gigliotti (1993).

6 Networked Minds and Distributed Cognition

I have stressed that there is one assumption in the derivation of the rational actor model that conflicts with the repeatedly observed fact that human minds are not isolated instruments of ratiocination, but rather are networked and cognition is distributed over this network. I propose here an analytical tool, based on a refinement of the rational actor model proposed by Gilboa and Schmeidler (2001), for representing distributed cognition. Following Gilboa and Schmeidler we assume there is a single decision-maker, say Alice, who faces a *problem* p such that each *action* a that Alice takes leads to some *result* r. Alice does not know the probability distribution of outcomes following action a, so she searches her memory for similar problems she has faced in the past, the action she has taken for each problem, and the result of her action. Thus her memory M consists of a set of *cases* of the form (q, a, r), where q is a problem, a is the action she took facing this problem, and r was the result of the action. Alice has a utility function $u(r)$ defined over results, and a *similarity function* $s(p, q)$ representing how "similar" her current problem p is to any past problem q that she has encountered. Gilboa and Schmeidler then present a set of plausible axioms that imply Alice will choose her action a to maximize the expression

$$\sum_{(q,a,r)\in M_a} s(p,q)u(r) \tag{11}$$

where M_a is the subset of Alice's memory where she took action a.

Several empirical studies have shown that this case-based decision approach is superior to other more standard approaches to choice under radical uncertainty (Gayer et al. 2007, Golosnoy and Okhrin 2008, Ossadnik et al. 2012, Guilfoos and Pape 2016). To extend this to distributed cognition, we simply replace Alice's personal memory bank by a wider selection of cases distributed over her social network of minds. It would also be plausible to add a second similarity function indicating how similar the individual who actually took the action is Alice herself.

7 Limitations of the Rational Actor Model

One often hears that a theory fails if there is a single counterexample. Indeed, this notion was the touchstone of Karl Popper's famous interpretation of the scientific method (Popper 2002[1959]). Because biological systems are inherently complex, this criterion is too strong for the behavioral sciences (Godfrey-Smith 2006, 2009; Weisberg 2007; Wimsatt 2007). Despite its general usefulness, the rational actor model fails to explain choice behavior in several well-known situations. Two examples are the famous Allais and Ellsberg Paradoxes. These are of course not paradoxes, but rather violations of rational choice.

7.1 The Allais Paradox

Maurice Allais (1953) offered the following scenario as a violation of rational choice behavior. There are two choice situations in a game with prizes $x =$ \$2,500,000, $y =$ \$500,000, and $z =$ \$0. The first is a choice between lotteries $\pi = y$ and $\pi' = 0.1x + 0.89y + 0.01z$. The second is a choice between $\rho = 0.11y + 0.89z$ and $\rho' = 0.1x + 0.9z$. Most people, when faced with these two choice situations, choose $\pi \succ \pi'$ and $\rho' \succ \rho$. Which would you choose?

This pair of choices is not consistent with the expected utility principle. To see this, let us write $u_h = u(2500000)$, $u_m = u(500000)$, and $u_l = u(0)$. Then if the expected utility principle holds, $\pi \succ \pi'$ implies $u_m > 0.1u_h + 0.89u_m + 0.01u_l$, so $0.11u_m > 0.10u_h + 0.01u_l$, which implies (adding $0.89u_l$ to both sides) $0.11u_m + 0.89u_l > 0.10u_h + 0.9u_l$, which says $\rho \succ \rho'$.

Why do people make this mistake? Perhaps because of *regret*, which does not mesh well with the expected utility principle (Loomes 1988, Sugden 1993). If you choose π' in the first case and you end up getting nothing, you will feel really foolish, whereas in the second case you are probably going to get nothing anyway (not your fault), so increasing the chances of getting nothing a tiny bit (0.01) gives you a good chance (0.10) of winning the really big prize. Or perhaps because of *loss aversion*, because in the first case, the anchor point (the most likely outcome) is \$500,000, while in the second case the anchor is \$0. Loss-averse individuals then shun π', which gives a positive probability of loss whereas in the second case, neither lottery involves a loss, from the standpoint of the most likely outcome.

The Allais paradox is an excellent illustration of problems that can arise when a lottery is consciously chosen by an act of will and one *knows* that one has made such a choice. The regret in the first case arises because if one chose the risky lottery and the payoff was zero, one knows for certain that one made a poor choice, at least ex post. In the second case, if one received a zero payoff, the odds are that it had nothing to do with one's choice. Hence, there is no regret in the second case.

But in the real world, most of the lotteries we experience are chosen by default, not by acts of will. Thus, if the outcome of such a lottery is poor, we feel bad because of the poor outcome but not because we made a poor choice.

7.2 The Ellsberg Paradox

Another classic violation of the expected utility principle was suggested by Daniel Ellsberg (1961). Consider two urns. Urn A has 51 red balls and 49 white balls. Urn B also has 100 red and white balls, but the fraction of red balls is unknown. One ball is chosen from each urn but remains hidden from sight. Subjects are asked to choose in two situations. First, a subject can choose the ball from urn A or urn B, and if the ball is red, the subject wins \$10. In the second situation, the subject can choose the ball from urn A or urn B, and if the ball is white, the subject wins \$10. Many subjects choose the ball from urn A in both cases. This violates the expected utility principle no matter what probability the subject places on the probability p that the ball from urn B is white. For in the first situation, the payoff from choosing urn A is $0.51u(10) + 0.49u(0)$ and the payoff from choosing urn B is $(1 - p)u(10) + pu(0)$, so strictly preferring urn A means $p > 0.49$. In the second situation, the payoff from choosing urn A is $0.49u(10) + 0.51u(0)$ and the payoff from choosing urn B is $pu(10) + (1 - p)u(0)$, so strictly preferring urn A means $p < 0.49$. This shows that the expected utility principle does not hold.

Whereas the other proposed anomalies of classical decision theory can be interpreted as the failure of linearity in probabilities, regret, loss aversion, and epistemological ambiguities, the Ellsberg paradox appears to strike even more deeply because it implies that humans systematically violate the following principle of first-order stochastic dominance (FOSD).

> Let $p(x)$ and $q(x)$ be the probabilities of winning x or more in lotteries A and B, respectively. If $p(x) \geq q(x)$ for all x, then $A \succeq B$.

The usual explanation of this behavior is that the subject *knows* the probabilities associated with the first urn, while the probabilities associated with the second urn are *unknown*, and hence there appears to be an added degree of risk associated with choosing from the second urn rather than the first. If decision-makers are risk-averse and if they perceive that the second urn is considerably riskier than the first, they will prefer the first urn. Of course, with some relatively sophisticated probability theory, we are assured that there is in fact no such additional risk, so it is hardly a failure of rationality for subjects to come to the opposite conclusion. The Ellsberg paradox is thus a case of performance error on the part of subjects rather than a failure of rationality.

7.3 Failures of Judgment

Contemporary behavioral economics has developed a powerful critique of the standard assumption that people are instrumentally rational (Ariely 2010, Thaler and Sunstein 2008). In fact, human decision-makers are close to instrumentally rational when they are sufficiently informed and the cost of exploring alternative strategies is low (Gintis 2009, Gigerenzer 2015). Nevertheless, the behavioral economics critique of the assumption of instrumental rationality is important and well-taken.

But as we have seen, the rational actor model depicts *formal rationality*, not *instrumental rationality*. That is, it assumes that people have consistent preferences and update according to Bayes rule, but it does not assume that rational behavior is oriented towards any particular end state or goal, and certainly not that rational behavior furthers the fitness or welfare interests of the decision-maker. Let us review the major claims made by behavioral economists supporting the notion that choice behavior is fundamentally irrational.

- *Logical Fallibility*: Even the most intelligent decision-makers are prone to commit elementary errors in logical reasoning. For example, in one well-known experiment performed by Tversky and Kahneman (1983), a young woman, Linda, is described as politically active in college and highly intelligent. The subject is then asked the relative likelihood of several descriptions of Linda, including the following two: "Linda is a bank teller" and "Linda is a bank teller and is active in the feminist movement." Many subjects rate the latter statement more likely than the former, despite the fact that the most elementary reasoning shows that if p implies q, then p cannot be more likely than q. Because the latter statement implies the former, it cannot be more likely than the former.

- *Anchoring*: When facing extreme uncertainty in making an empirical judgment, people often condition their behavior on recent but irrelevant experience. For instance, suppose a subject is asked to write down a number equal to the last two digits of his social security number and then to consider whether he would pay this number of dollars for particular items of unknown value. If he is then asked to bid for these items, he is likely to bid more if the number he wrote down was higher.

- *Cognitive Bias*: If you ask someone to estimate the result of multiplying $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$, he is likely to offer a lower estimate than if you had presented him with $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$. Similarly, if you ask someone what fraction of English words end in "ng" and give the example "gong," you will probably get a lower estimate than if you gave the example "going."

- *Availability Heuristic*: People tend to predict the frequency of an event based on how often they have heard about it. For example, most people believe that homicides occur with more frequency than suicides, although the reverse is the case. Similarly, they believe that certain cancers cluster in certain communities because of environmental pollutants, where in fact such clusters may occur no more frequently than chance, but are more likely to be reported.

- *Status Quo Bias*: Decision-makers tend to follow a certain traditional pattern of behaviors even after there is strong credible evidence that a superior course of action is available. For instance, in an early well-known experiment, Samuelson and Zeckhauser (1988) presented subjects with a task in which several financial assets were listed and the subjects were asked to choose one that they prefer to invest in. A second set of subjects was given the same list of financial assets, but one was presented as the *status quo*. They found that the asset listed as the *status quo* was chosen at a much higher frequency than when it was presented just as one among several randomly presented alternatives.

- *Herd Mentality*: People are heavily influenced by the actions of others. For instance, Solomon Asch (1951) showed that peer pressure can induce subjects to offer clearly false evaluations, even when the subject and his peers do not know each other and will likely never meet outside the laboratory. Groups were formed consisting of eight college students, all but one of whom were confederates of the experimenter. Each student was shown a card with a black line on it, and a second card with three black lines, one of which was the same length as the line on the first card, and the other two were of very different lengths, one longer and the other shorter. Each student was asked to say out loud which line on the second card matched the line on the first card, the seven confederates going first and choosing an obviously incorrect line. More than a third of subjects agreed with the obviously wrong answer.

- *Framing Effects*: A framing effect is a form of cognitive bias that occurs when choice behavior depends on the wording of two logically equivalent statements. Take, for example, the classic example of the physician and his heart patient, analyzed by McKenzie et al. (2006) and Thaler and Sunstein (2008). The patient must decide whether to have heart surgery or not. His doctor tells him either (A) “Five years after surgery, 90% of patients are alive,” or (B) “Five years after surgery, 10% of patients are dead.” The two statements are of course logically equivalent, but subjects are far more likely to accept surgery with frame (A) than with frame (B).

- *Default Effects*: In choosing among various options, if one is offered as the default option, people tend to choose it with high frequency. A most dramatic example is organ donation (Johnson and Goldstein 2003). Countries in Europe that have a presumed consent default have organ donation rates that are about 60% higher than countries with explicit consent requirements. Another famous example involves registering new employees in a company 401(k) savings plan. When participating is the default, participation is considerably higher than when the default is non-participation (Bernheim et al. 2011).

The logical fallibility argument would of course be devastating to rational choice theory, which implicitly assumes that decision-makers are capable of making logical deductions. There are certainly complex logical arguments that the untrained subject is likely to get wrong. Indeed, even a mistake in mathematical computation counts as an error in logical reasoning. But what appear to be the elementary errors of the type revealed by the Linda the Bank Teller example are more likely to be errors of interpretation on the part of the experimenters. It is important to note that given the description of Linda, the probability that an individual is Linda if we know that the individual is a bank teller is much lower than the probability that an individual is Linda if we know that she is a feminist bank teller. This is because Linda is probably a feminist, and there are far fewer feminist bank tellers than there are bank tellers. Subjects in the experiment might reasonably assume that the experimenters were looking for a conditional probability response rather than a simple probability response because they supplied a mass of information that is relevant to conditional probability, but is quite irrelevant to simple probability.

Indeed, in normal human discourse, a listener assumes that any information provided by the speaker is relevant to the speaker’s message (Grice 1975). Applied to this case, the norms of discourse reasonably lead the subject to believe that the experimenter wants Linda’s politically active past to be taken adequately into account (Hilton 1995, Wetherick 1995). Moreover, the meaning of such terms as “more likely” or “higher probability” are vigorously disputed even in the theoretical literature, and hence are likely to have a different meaning for the average subject versus for the expert. For example, if I were given two piles of identity folders and asked to search through them to find the one belonging to Linda, and one of the piles was “all bank tellers” while the other was “all bank tellers who are active in the feminist movement,” I would surely look through the latter (doubtless much smaller) pile first, even though I am well aware that there is a "higher probability" that Linda's folder is in the former pile rather than the latter one. In other words, conditional rather than straight probability is the appropriate concept in this case.

However important anchoring, cognitive bias, and the availability heuristic may be, they are clearly not in conflict with the rational actor model because they do not compromise any of the rational choice axioms. In particular, they do not imply preference inconsistency or the failure of Bayesian updating. The *status quo* bias may seem to contradict Bayesian updating, but it does not. For one thing, if one is satisfied with a particular choice, it may plausibly appear excessively costly to evaluate properly new information. Herbert Simon (1972) called this reasonable behavior "satisficing." It is clearly compatible with rational updating. For another, one may reasonably ignore new information on the grounds that it is unreliable. Models of Bayesian updating simply assume that the new information is rigorously factual, which is often not the case.

The framing effects literature is more challenging. Indeed, some argue that because it is impossible to avoid framing effects, there are no true underlying preferences, so the rational actor model fails. This conclusion is unwarranted. In this book we specified from the outset that preferences are generally state-, time-, and social frame-dependent. In particular, preferences are frame-dependent because individual choices, except perhaps for Robinson Crusoe before he meets Friday, occur within a social context, and that context is the social frame for choice behavior. Indeed, even the absence of a social frame is a social frame.

Consider, for instance, the physician and his heart patient scenario described above. Thaler and Sunstein (2008) interpret this as showing that many decision-makers are irrational. But it is more accurate to interpret these results as patients simply following the implicit suggestion of the physician, the expert on whom their well-being depends. Note first that neither (A) nor (B) gives the patient sufficient information to make an informed choice because the physician does not provide the equivalent survival and death rates *without* surgery. The only reasonable inference is that the patient believes the doctor is recommending surgery in case (A), and recommending against surgery in case (B).

The physician and his heart patient example is not an isolated case of the tendency for behavioral economists to ignore the intimately social nature of choice, and to interpret completely reasonable behavior as irrational. Gigerenzer (2015), who documents several additional examples of this tendency, concludes:

> Research...indicates that logical equivalence is a poor general norm for understanding human rationality...Speakers rely on framing in order to implicitly convey relevant information and make recommenda-

> tions, and listeners pay attention to these. In these situations, framing effects clearly do not demonstrate that people are mindless, passive decision-makers.

Similarly, default effects do not illustrate the decision-maker's irrationality, but rather the tendency to treat the default as a recommendation by experts whose advice it is prudent to follow unless there is good information that the default is not the best choice (Johnson and Goldstein 2003). Indeed, Gigerenzer (2015) reports that a systematic review of hundreds of framing studies could not find a single one showing that framing effects incur real costs in terms.

REFERENCES

Ahlbrecht, Martin and Martin Weber, "Hyperbolic Discounting Models in Prescriptive Theory of Intertemporal Choice," *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 115 (1995):535–568.

Allais, Maurice, "Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école Américaine," *Econometrica* 21 (1953):503–546.

Andreoni, James, "Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments," *Quarterly Journal of Economics* 110,1 (February 1995):1–21.

Ariely, Dan, *Predictibly Irrational:The Hidden Forces That Shape Our Decisions* (New York: Harper, 2010).

Asch, Solomon, "Effects of Group Pressure on the Modification and Distortion of Judgments," in H. Guetzkow (ed.) *Groups, Leadership and Men* (Carnegie Press, 1951) pp. 177–190.

Benabou, Roland and Jean Tirole, "Self Confidence and Personal Motivation," *Quarterly Journal of Economics* 117,3 (2002):871–915.

Berg, Joyce E., John W. Dickhaut, and Thomas A. Rietz, "Preference Reversals: The Impact of Truth-Revealing Incentives," 2005. College of Business, University of Iowa.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov, "The Welfare Economics of Default Options in 401 (k) Plans," 2011. National Bureau of Economic Research, No. w17587.

Brainard, D. H. and W. T. Freeman, "Bayesian Color Constancy," *Journal of the Optical Society of America A* 14 (1997):1393–1411.

Brewer, Marilyn B. and Roderick M. Kramer, "Choice Behavior in Social Dilemmas: Effects of Social Identity, Group Size, and Decision Framing," *Journal of*

Personality and Social Psychology 50,543 (1986):543–549.

Carey, Susan, *Conceptual Change in Childhood* (Cambridge: MIT Press, 1985).

Carpenter, Jeffrey P., Stephen V. Burks, and Eric Verhoogen, "Comparing Student Workers: The Effects of Social Framing on Behavior in Distribution Games," *Research in Experimental Economics* 1 (2005):261–290.

Cookson, R., "Framing Effects in Public Goods Experiments," *Experimental Economics* 3 (2000):55–79.

Durrett, Richard and Simon A. Levin, "Can Stable Social Groups be Maintained by Homophilous Imitation Alone?," *Journal of Economic Behavior and Organization* 57,3 (2005):267–286.

Ellsberg, Daniel, "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics* 75 (1961):643–649.

Fischer, Ilan, Alex Frid, Sebastian J. Goerg, Simon A. Levin, Daniel I. Rubenstein, and Reinhard Selten, "Fusing Enacted and Expected Mimicry Generates a Winning Strategy that Promotes the Evolution of Cooperation," *Proceedings of the National Academy of Sciences* 110,25 (June 18 2013):10229–10233.

Gayer, G., Itzhak Gilboa, and O. Lieberman, "Rule-based and Case-based Reasoning in Housing Prices," *The BE Journal of Theoretical Economics* 7,1 (2007).

Gigerenzer, Gerd, "On the Supposed Evidence for Libertarian Paternalism," *Review of Philosophical Psychology* 6 (2015):361–383.

— and P. M. Todd, *Simple Heuristics That Make Us Smart* (New York: Oxford University Press, 1999).

Gilboa, Itzhak and David Schmeidler, *A Theory of Case-Based Decisions* (Cambridge: Cambridge University Press, 2001).

Gintis, Herbert, "A Framework for the Unification of the Behavioral Sciences," *Behavioral and Brain Sciences* 30,1 (2007):1–61.

— , *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences* (Princeton: Princeton University Press, 2009).

— , *Individuality and Entanglement: The Moral and Material Bases of Human Social Life* (Princeton: Princeton University Press, 2016).

Glymour, Alison, D. M. Sobel, L. Schultz, and C. Glymour, "Causal Learning Mechanism in Very Young Children: Two- Three- and four-year-olds Infer Causal Relations from Patterns of Variation and Covariation," *Developmental Psychology* 37,50 (2001):620–629.

Glymour, C., *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology* (Cambridge: MIT Press, 2001).

Godfrey-Smith, Peter, "The Strategy of Model-Based Science," *Biology and Philosophy* 21 (2006):725–740.

— , "Models and Fictions in Science," *Philosophical Studies* 143 (2009):101–116.

Golosnoy, V. and Y. Okhrin, "General Uncertainty in Portfolio Selection: A Case-based Decision Approach," *Journal of Economic Behavior and Organization* 67,3 (2008):718–734.

Gopnik, Alison and Andrew Meltzoff, *Words, Thoughts, and Theories* (Cambridge: MIT Press, 1997).

— and Joshua B. Tenenbaum, "Bayesian Networks, Bayesian Learning and Cognitive Development," *Developmental Studies* 10,3 (2007):281–287.

— and L. Schultz, *Causal Learning, Psychology, Philosophy, and Computation* (Oxford: Oxford University Press, 2007).

Grether, David and Charles R. Plott, "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 69,4 (September 1979):623–638.

Grice, H. P., "Logic and Conversation," in Donald Davidson and Gilbert Harman (eds.) *The Logic of Grammar* (Encino, CA: Dickenson, 1975) pp. 64–75.

Guilfoos, Todd and Andreas Duus Pape, "Predicting Human Cooperation in the Prisoner's Dilemma Using Case-based Decision Theory," *Theory and Decision* 80 (2016):1–32.

Haidt, Jonathan, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon, 2012).

Hilton, Denis J., "The Social Context of Reasoning: Conversational Inference and Rational Judgment," *Psychological Bulletin* 118,2 (1995):248–271.

Johnson, Eric J. and Daniel G. Goldstein, "Do Defaults Save Lives?," *Science* 302 (2003):1338–1339.

Kirby, Kris N. and Richard J. Herrnstein, "Preference Reversals Due to Myopic Discounting of Delayed Reward," *Psychological Science* 6,2 (March 1995):83–89.

Knill, D. and A. Pouget, "The Bayesian Brain: the Role of Uncertainty in Neural Coding and Computation," *Trends in Cognitive Psychology* 27,12 (2004):712–719.

Lichtenstein, Sarah and Paul Slovic, "Reversals of Preferences between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology* 89 (1971):46–55.

Loomes, Graham, "When Actions Speak Louder than Prospects," *American Economic Review* 78,3 (June 1988):463–470.

Luce, Robert Duncan, *Individual Choice Behavior* (New York: Dover, 2005).

— and Patrick Suppes, "Preference, Utility, and Subjective Probability," in Robert

Duncan Luce, Robert R. Bush, and Eugene Galanter (eds.) *Handbook of Mathematical Psychology, vol. III* (New York: Wiley, 1965).

Machina, Mark J., "Choice under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives* 1,1 (Summer 1987):121–154.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in Paul Zarembka (ed.) *Frontiers in Econometrics* (New York: Academic Press, 1973) pp. 105–142.

McKenzie, C. R. M., M. J. Liersch, and S. R. Finkelstein, "Recommendations Implicit in Policy Defaults," *Psychological Science* 17 (2006):414–420.

McPherson, M., L. Smith-Lovin, and J. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* 27 (2001):415–444.

Ok, Efe A. and Yusufcan Masatlioglu, "A General Theory of Time Preference," 2003. Economics Department, New York University.

Ossadnik, W., D. Wilmsmann, and B. Niemann, "Experimental Evidence on Case-based Decision Theory," *Theory and Decision (*2012):1–22.

Pearl, J., *Causality* (New York: Oxford University Press, 2000).

Popper, Karl, *The Logic of Scientific Discovery* (London: Routledge Classics, 2002[1959]).

Risen, Jane L., "Believing what we do not Believe," *Psychological Review (*October 2015):1–27.

Samuelson, William and Richard Zeckhauser, "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty* 1 (1988):7–59.

Savage, Leonard J., *The Foundations of Statistics* (New York: John Wiley & Sons, 1954).

Schultz, L. and Alison Gopnik, "Causal Learning across Domains," *Developmental Psychology* 40 (2004):162–176.

Simon, Herbert, "Theories of Bounded Rationality," in C. B. McGuire and Roy Radner (eds.) *Decision and Organization* (New York: American Elsevier, 1972) pp. 161–176.

Sobel, D. M. and N. Z. Kirkham, "Bayes Nets and Babies: Infants' Developing Statistical Reasoning Abilities and their Representations of Causal Knowledge," *Developmental Science* 10,3 (2007):298–306.

Sopher, Barry and Gary Gigliotti, "Intransitive Cycles: Rational Choice or Random Error: An Answer Based on Estimation of Error Rates with Experimental Data," *Theory and Decision* 35 (1993):311–336.

Spirtes, P., C. Glymour, and R. Scheines, *Causation, Prediction, and Search* (Cambridge: MIT Press, 2001).

Stalnaker, Robert, "A Theory of Conditionals," in Nicholas Rescher (ed.) *Studies in Logical Theory* (London: Blackwell, 1968).

— , "Knowledge, Belief, and Counterfactual Reasoning in Games," *Economics and Philosophy* 12 (1996):133–163.

Starmer, Chris, "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature* 38 (June 2000):332–382.

Steyvers, Mark, Thomas L. Griffiths, and Simon Dennis, "Probabilistic Inference in Human Semantic Memory," *Trends in Cognitive Sciences* 10 (2006):327–334.

Sugden, Robert, "An Axiomatic Foundation for Regret Theory," *Journal of Economic Theory* 60,1 (June 1993):159–180.

Thaler, Richard H. and Cass Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (New York: Penguin, 2008).

Tooby, John and Leda Cosmides, "The Psychological Foundations of Culture," in Jerome H. Barkow, Leda Cosmides, and John Tooby (eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (New York: Oxford University Press, 1992) pp. 19–136.

Tversky, Amos and Daniel Kahneman, "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement," *Psychological Review* 90 (1983):293–315.

— , Paul Slovic, and Daniel Kahneman, "The Causes of Preference Reversal," *American Economic Review* 80,1 (March 1990):204–217.

von Neumann, John and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton: Princeton University Press, 1944).

Weisberg, Michael, "Who is a Modeler?," *British Journal of the Philosophy of Science* 58 (2007):207–233.

Wetherick, N. E., "Reasoning and Rationality: A Critique of Some Experimental Paradigms," *Theory & Psychology* 5,3 (1995):429–448.

Wimsatt, William C., *Re-Engineering Philosophy for Limited Beings* (Cambridge: Harvard University Press, 2007).

c\papers\\.tex March 6, 2017