

Computable Approaches to Rational Choice and Decision-Making

Hyunjin Seo
 University of Kansas
hseo@ku.edu

Stuart Thorson
 Syracuse University
thorson@syr.edu

Abstract

The capacity of agents to act rationally, that is to make choices that positively reflect their interests, is a core assumption underlying democratic governance systems, microeconomics, decision science, market driven economies, and many agent based modeling efforts. In this paper we investigate axiomatic theories of rational choice from the perspective of computability. Using algorithmic complexity, we show highly general conditions under which no effective procedure can exist enabling these theories to identify sequences of choices as random. While axiomatic theories of rational choice yield powerful descriptions of choice behavior, this power comes at the expense of axioms which can be brittle with regard to computability limits.

1. Introduction

To say of someone that they are rational is to describe them in a particular way and it is reasonable to ask what are the consequences of such a description. The purpose of this paper is to explore this question from the perspective of computability. Whether or not humans are subject to fundamental computational limits, it is our argument that scientific theories of agents as rational choosers depend on an assumption that people can be usefully described as entities whose choices are calculated and thus subject to computational constraints. Rational choice assumptions are common in areas including decision science, microeconomics, voting behavior, and agent-based modeling.

Within the axiomatic or formal theory of rationality tradition¹ there is a broadly shared sense that *rationality*

¹ Here we refer to theories of rationality which state explicit axioms about rational choice and use these to derive statements about what we should expect from agents whose behaviors satisfy those axioms.

requires some sort of positive responsiveness between underlying preferences and visible choices [1]. When offered a pair of items from which to make a choice, a rational agent will choose the one she most prefers from the two.

A basic intuition is that a choice by a rational agent is a choice that can be made sense of by knowing the agent's underlying preferences. However, it is precisely these underlying preferences which we, as external observers, generally cannot directly measure. We must infer them from choices we can observe. Any particular axiomatic theory of rational choice will then tell us something about what to expect about choice behaviors given inferred information about underlying preferences. A theory of rational choice should provide informative predictions regarding a rational agent's choice behavior.

The literature contains many examples of rigorously formulated rational choice theories. With a notable exception being those influenced by Herbert Simon [2], these rarely address issues of a person's actual capacity to make choices in accord with the theory. While they identify formal conditions under which a rational choice can exist, they do not provide rules for actually computing that choice [3].

A key insight of Simon's was that complexity can affect a person's choice behavior and sometimes preclude making a globally best choice. In selecting a life partner there are billions of possible alternatives from which to select. However, a person will have relevant information about only a subset of these. Additionally, the final choice may be influenced by prior considerations of alternatives. The result is a *satisficing* or good enough selection exemplifying Simon's notion of *bounded rationality*.

Simon emphasizes processing limitations on human choices resulting from complex information environ-

ments such as where the number of alternatives is large. As further developed by Selten [4], this fundamentally involves empirical questions. In contrast, following [5]–[8] we aim to identify basic computability constraints on assessing the rationality of choices.

We proceed by noting that axiomatic theories can be understood or interpreted as computer programs, or equivalently as Turing machines. These theories are intended to produce (permit derivations of) descriptions that are informative with regard to rational choice behavior. We should be able to provide a formal theory of rationality some input, say preferences and a set of alternatives from which a choice is to be made, and have the theory produce as output descriptions of the choice an agent satisfying the axioms of the theory would make.

Moreover, not just any such theory would do. A theory which takes as input the choices made by an agent and then reproduces as output descriptions of those choices would add no new information. We expect rational choice axioms together with a set of inference rules to offer a compressed account of patterns of choice behavior in a manner analogous to how a PNG image together with an appropriate decompression program provides a compressed description of a photograph. This expectation is grounded on the intuition that an agent’s choices, if rational, should exhibit some sort of patterning. That is, they should not be algorithmically random. The particular patterning will depend upon the axioms of the theory being used. Using algorithmic complexity we show general conditions under which no effective procedure can exist to enable these theories to identify sequences of choices as maximally random. While axiomatic theories of rational choice yield powerful descriptions of choice behavior, this power comes at the expense of axioms which can be brittle with regard to computability limits.

2. Axiomatic Theories of Rationality and Computability

Theories of rationality discussed here are those consisting of axioms which, together with fairly basic arithmetic and logic operators, permit derivation of statements describing observable patterns of behavior by a rational chooser. We abbreviate class of these axiomatic theories of rational choice as *ATRC*. While there are

a variety of axiom sets within *ATRC*, at core they involve describing what happens when a rational chooser is faced with a number of options from which some sort of choice is to be made. We will assume the consistency of theories within *ATRC*.

Our interest is in the axiomatic structure of the class of formal theories of rational choice once computability condition is required. These theories typically begin by positing the existence of a set of alternatives, X , from which an agent is to make a choice. A preference relation is a binary relation on the Cartesian product of the alternative set, $X \times X$. Typical conditions are those of the strict preference relation, P defined to be:

- i Irreflexive: $\forall x \in X, \neg(xPx)$,
- ii Asymmetric: $\forall x, y \in X, \neg(xPy \wedge yPx)$,
- iii Transitive: $\forall x, y, z \in X, xPy \wedge yPz \implies xPz$, and
- iv Complete: $\forall x, y \in X, xPy \vee yPx \vee y = x$.

P then yields a strict total order on X .

Some axiomatic theories of rationality consider the choice process as deterministic while others see it as probabilistic. Then there are theories of individual choice behavior [9] and those that focus on the behaviors of groups or aggregates. In this paper we consider theories of single-agent rationality where rational choices are assumed to flow in a positive fashion from underlying preferences. Our interest is in the relation between preferences and choices and we will not be concerned with numeric representations of preferences via utility functions²

Our example formalization of *choice* is patterned after the fundamental work of Uzawa [11], Arrow [12], and Richter [13] but most other formalizations should work as well. We begin with an agent faced with making a choice. The objective is to establish conditions for deciding whether or not the agent’s choice behavior can be described as *rational*. Consider a *choice space* $\langle X, \mathcal{B} \rangle$ where X represents a nonempty set of alternatives possibly available for choice and \mathcal{B} is comprised of nonempty subsets of X where each $B \in \mathcal{B}$ corresponds to a set of alternatives actually available for choice on a given occasion.³

²For utility representations see [3], [10].

³Some theories require X to be finite and others assume it is denumerably infinite or even uncountably infinite (e.g., [13, p. 635] considers the case where X is represented by the “nonnegative orthant of the n -fold Cartesian product of the real line”). We require X to be countable.

Computational complexity issues arise depending upon, for instance, whether \mathcal{B} is assumed to include all nonempty subsets of X . In such a case, \mathcal{B} would grow exponentially with the size of X according to the formula $2^n - 1$ where n is the number of alternatives in X . For $n = 3$, the size of \mathcal{B} would be only seven. However, if there are 20 alternatives, \mathcal{B} grows to 1,048,575.

In an economics context, a B might be the various bundles of goods affordable given a budget. In a voting context, a B might be a particular ballot consisting of names of candidates from which to choose. We shall refer to the pair $\langle X, \mathcal{B} \rangle$ as a *ballot space* with the understanding that the particular interpretation will depend upon the contents of X . So if X consists of candidates for office, \mathcal{B} would be subsets of candidates and a particular element of \mathcal{B} would describe a possible ballot. A choice function is a function h which for every $B \in \mathcal{B}$ yields a nonempty subset $h(B) \subseteq B$ representing the selection or choice made from each B in \mathcal{B} [13].

Our focus is on preference-based individual choice and we define a *strict preference* binary relation P on X as above such that P is irreflexive, asymmetric, transitive, and total.⁴ The intended interpretation here is that P is a preference relation such that for any (x_i, x_j) where $i \neq j$ in X either $(x_i P x_j)$ or $(x_j P x_i)$. If $x_i P x_j$ we say x_i preferred to x_j .

Since P is a total and transitive binary relation, choices may reveal information about preferences even though those preferences are not directly observable. An agent can be described as choosing rationally if, given a set of alternatives, the choice is the P greatest one. Formally, a P greatest choice is one which can be described by a choice function h defined on \mathcal{B} such that:

$$\forall B \in \mathcal{B}, h(B) = x^* : x^* \in B \wedge \forall (x_i \neq x^*) \in B, x^* P x_i.$$

This account of a best choice tells us that to describe an agent as choosing rationally is to say that the choice can be described by an h that rationalizes it. But how are we to know whether an agent's choice is, in empirical fact, rational? One approach would be to first suppose a person to be rational and then note that each time a

⁴The reason for requiring X to be denumerable is now apparent. If X were uncountably infinite as with the real line, then how could we define an effective procedure for determining for every distinct pair in X whether or not it was in P ? [8] provides an extended discussion of this.

person chooses an alternative from a B she does reveal something about her preferences. Imagine we observe a person's choice behaviors over the full range of B s while keeping h fixed. If a chooser selects alternative x_i from a B when, say, x_j and x_k are also in B we can say that x_i has been *revealed* as being preferred to both x_j and x_k . This then leads to identifying a rational agent as one whose observable choice behavior satisfies the principle that for every possible B in \mathcal{B} her choice consists of the P greatest alternative.

A concrete example, though tedious, will clarify the preceding discussion and help set up what follows. Consider a case in which there are three brands of toothpaste being considered for purchase—AIM, BABOOL, and CREST represented using the first letter of their name: $X = \{a, b, c\}$. The maximal size of any \mathcal{B} is $2^n - 1$ (the total number of possible subsets of a set with n distinct elements is 2^n and from that we subtract 1 since the definition of \mathcal{B} stipulates that it contains only nonempty subsets) yielding, in this instance 7. So the largest possible \mathcal{B} would consist of all nonempty subsets of X or $\mathcal{B} = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$.

Suppose a person's preferences happen to order the possible purchases alphabetically:

$P = \{(a, b), (a, c), (b, c)\}$. P is irreflexive, asymmetric, transitive, and total. If the person chooses rationally we would expect her choice function to look as follows: $h(\{a\}) \rightarrow a, h(\{b\}) \rightarrow b, h(\{c\}) \rightarrow c, h(\{a, b\}) \rightarrow a, h(\{a, c\}) \rightarrow a, h(\{b, c\}) \rightarrow b$, and $h(\{a, b, c\}) \rightarrow a$. It is straightforward to verify that the preferences revealed by these choices do indeed rationalize the choices indicated by h .

This example can be modified to construct an example of an irrational choice. Consider a function h^* everywhere identical to h above except that $h^*(\{b, c\}) \rightarrow c$. It is simple to verify that there is no total, transitive and strongly reflexive P^* that could rationalize those choices and thus an agent whose choices were those of h^* would not be describable as making everywhere rational choices. This demonstrates the logical possibility of an irrational choice.

The logical possibility of a person making an irrational choice suggests that rationality is an assumption and not merely a tautology asserting anything a person chooses to be rational by virtue of its having been chosen when some other alternative could have been selected. This is significant as rationality arguments are often criticized as shape shifting when shown that an

agent's choices deviate from what a particular rational choice theory predicts. A response to the faulty prediction might be that, in this particular context, P ought really be thought of as probabilistic relation and we should be speaking of, say, the probability that a person will select chicken over beef when at a restaurant rather than requiring a deterministic chicken preference over beef. In this paper we will sidestep debates as to precisely which axioms are best suited to describing rationality and focus on computational constraints relevant to the class of axiomatic rational choice arguments.

A common thread in *ATRC* is that a person's choices are not random but rather reveal something about the structure and, perhaps, content of their underlying preferences at the time of their choice. Additionally, the context or environment in which a choice is made may matter. For example, in a context of risk, probabilistic assumptions about P may make sense. While in determining how to fill out election ballots, viewing P as deterministic may be compelling.⁵ In other words, if we consider a person's choice behavior over time it is reasonable to think that a person may deploy different procedures for making choices depending upon conditions at the time.

Therefore to determine whether a person's choices can be said to be rational in both a single choice situation and when looking at that person's choices over time we must go beyond asking whether a particular choice is rational and decide whether the choices a person makes tend to positively reflect their preferences. Here we recognize that a choice which is rational under one set of *ATRC* assumptions might appear not so under a different set. Further, there is the possibility that, on occasion, a person, from her own perspective, errs in making a choice. A particular choice may result from intending to do one thing yet actually doing another (as in intending to vote for candidate A but mistakenly pulling the lever for candidate B) or from doing what we intended but having that doing' incorrectly recorded (as in properly casting a ballot for A on a flawed voting machine).

As a person lives her life, she makes a sequence of choices, $\{c_1, c_2, \dots, c_n\}$. A person can be described as rational if that string of choices tends to positively re-

⁵A claim that different contexts require different axioms may be seen begging the question unless accompanied by guidance as to how an agent should rationally determine which axioms to use when. But then must not that guidance also be justified? And so on.

flect her underlying preferences. For that to be the case, it should be her choices exhibit patterning suggesting they can be rationalized by some, though not necessarily everywhere the same, choice function. By *patterned* here we mean that the relation between preferences and choices is not random. Equivalently, knowing a person's preferences over time should enable us to make better than chance predictions about her choice behavior.

When being rational, a person acts with intention. She intends her choices to positively reflect her preferences. Second, as with most intentions, they can be overridden by circumstance and for that reason we cannot anticipate every c_i for a rational person to be rationalized by her preferences. For example, it is not uncommon to make a purchasing decision and later, upon reflection, conclude that our preferences would have led us to a different choice.

We take as a given that to model a person's choices computationally is to assume that an agent can be simulated, to any desired precision, by a mechanism that computes. By computation we mean a discrete state mechanism of unbounded size capable of reading from a tape, writing on that tape, and executing loops and conditional branches. Such a machine, a Turing machine (*TM*), defines the class of recursively enumerable languages about which anything can be computed. To say that a function is computable is equivalent to saying the function can be executed on a suitable coded *TM*. This is the Church-Turing thesis and it is from this perspective that we investigate consequences of modeling rationality computationally; that is simulating rational choice as calculations made by an appropriately programmed Turing machine. We assume what Aaronson's terms the physical Church-Turing thesis that any device that exists in the physical world, if it behaves in a predictable way at all, can be simulated, given input data by a *TM* machine to any desired precision [14].

TMs are quite general and a given *TM* can be encoded and simulated on another *TM*. With suitable encoding, data and programs become interchangeable. By convention the encoding will be in terms of bit strings, strings of 0s and 1s. Any particular program can be encoded without loss as a string of bits as can numbers, photographs, or this paper. This permits reasoning about what is and is not computable and, given the Church-Turing thesis, what we can computationally simulate.

Of particular interest is the fact that any given *TM* can be uniquely described by the binary string that defines it. This gives rise to the notion of a universal Turing machine which takes the binary string description of any particular *TM* together with binary string input for that *TM* and then computes the same output string as would the *TM*. An example of this sort of thing is provided by programs emulating Windows machines under the MacOS operating system enabling using programs (that is, strings of bits) originally compiled for Windows under MacOS (another string of bits). Programs and data can each be encoded as binary strings.

Computability immediately imposes conditions that, while perhaps initially unfamiliar, are quite compatible with our desire to focus on what a chooser could, from a logical perspective, possibly do. In particular, there are problems which are known to be not computable. A well known example is the *halting problem*. Turing proved [15] that in general determining in advance from the description of a computer program whether that program will finish running (halts) or run forever is not a computable problem. There can be no effective procedure for taking any string of bits representing a computer program and deciding whether that program will halt or run forever.⁶ Beyond this, it is also known that there are classes of problems with solutions that are computable but nonetheless are practically intractable in the sense of taking a very long time to actually compute the results.

Our interest is in the computability of rational choices. This requires that alternatives not be represented by real numbers as the set of real numbers is uncountably infinite. Indeed, the computable reals, a subset of the real numbers, do not form a computable field as no effective equality relation can be defined as was proven by Rice [16]. This has particular importance given our focus on rationality, and provides a formal reason we focus on the strict preference relation, *P* [7].

We will now use this to identify a computational limitation to testing for rationality. Imagine we had a Laplacian Demon [17] able to calculate a person's preferences and choices. This would mean Demon could also identify which choices positively reflected preferences and thus were *rational* in the sense we have been

⁶Of course, such a decision can be made for some programs. For there to be an effective procedure, it must yield a yes/no decision for all of the countable set of computer programs.

discussing. Demon will have to compute answers to counterfactual questions of the sort "what would the person prefer if presented with the set $\{x_i, x_j\}$?" for all distinct pairs in the alternative set in order to calculate *P*.

Here a strictly rational chooser would be expected to choose the most preferred of the two alternatives for that would positively reflect the preference for that pair.⁷ However, as noted above, even a rational person may on occasion make mistakes or there may be exogenous circumstances affecting a choice. Demon, of course, can identify these choice "deviations" and properly categorize them as not rational without regard to the cause or reason.

Demon encodes a rational choice over a pair with a 1 and a choice that is not rational with a 0. A person's possible choices over pairs in the alternative set could then be encoded as a string of 1s and 0s. So, for example, if Demon looked at a person making choices over a three alternative set as in the AXP, BA, and CAG example it would calculate what the person would do over each of the three possible paired comparisons, $\{\{a, b\}, \{a, c\}, \{b, c\}\}$. If the person chose rationally in every instance we would have the string "111". Or if the person made a non rational choice over the third pair, the string might be "110".

As the number of alternatives increases, the length of the bit string increases as well. Suppose there are ten brands of toothpaste under active consideration for purchase. The number of distinct pairs with ten alternatives is 45 and the number of unique bit strings of length 45 is 2^{45} or 35,184,372,088,832. A fairly large number. A recent check on Wikipedia indicated about there were about 50 brands of toothpaste. Here the number of distinct pairs is 1,225 and the number of unique bit strings is $\sim 5.8 \times 10^{368}$. Compare this the estimated number of atoms in the universe, $\sim 10^{80}$.

This poses no particular computation problem to either the chooser or Demon as the number of paired comparisons to be made for small *n* alternative sets is modest and there do exist time efficient algorithms for doing paired comparisons. However, issues do arise when we consider the possibility of *deviations* in choices over paired comparisons. Even for a person intending to make rational choices over all the pairs, mistakes or er-

⁷Of course what that choice would be expected to be would depend upon the particular *ATRC*.

rors, might occur. Consequently, rather than Demon calculating a bit string of all 1s, “1, 1, 1, . . . 1, 1”, for a person intending to be rational, the actual calculated string may look like “1, 0, 1, 1, 0, . . . , 1, 0”. Suppose then that Demon provides us the bit string and invites us to decide whether it encodes the choices of a rational chooser. Here we do have to confront the combinatorial explosion of possible bit strings.

A way to proceed would be to note that a string from a rational agent, even one who makes many mistakes, will not be random and should exhibit some patterning. This can be used to compress the string without loss of information in much the same way the bit string describing pixels of a photograph can be compressed. A discrepancy-free rational person’s string would simply be a string of 1s as long as the number of choices. On the other hand, if a person’s choices are fully non rational, we would expect no structure as the person would be simply selecting alternatives with no intent beyond making a selection. This is not to say the chooser has no preferences, only that the choices made were not driven by them. Demon, as stipulated, calculates preferences and can accurately predict choices so can readily categorize a given choice as 1 or 0. Since paired comparisons are binary choices and we are assuming people have preferences, there is always the possibility that acting without intention, that is ignoring preferences, will still produce a rational selection.

However, we do not have a probability distribution of choice strings from which we are sampling. We have only the single string.

Classical probability theory provides guides for calculating the likelihood of particular outcomes given that those outcomes are generated by a stochastic process with known, or assumed to be known, properties. This is not helpful in our case as we have a single binary string and no particular warrant for assuming the existence of an underlying generating distribution.

This leads to looking at a given string in terms of its *algorithmic complexity* [18]. The basic idea is that some long strings can be described by simple algorithms. For example, we might be interested in the string representing the first trillion digits of π . A relatively short computer program can compute these digits and the same program is also capable of computing only the first million digits of π . Moreover, that program would be considerably shorter than the string of digits it was computing. In this sense, the first trillion digits of π is highly

compressible; it can be described by a much shorter string (the computer program that computes it plus overhead to generate that string).

Our interest is in a binary string provided by Demon describing a sequence of choices. Each such string uniquely describes a sequence of binary choices under the encoding outlined above. For sake of discussion, suppose we have some particular *ATRC* and we accept that the theory is both consistent and sound.

Let’s further assume that our *ATRC* contains enough arithmetic to do simple operations. Given a binary string describing choices, s , we can conceptualize the algorithmic complexity of s , $K(s)$, as the size of the smallest string consistent with our *ATRC* that, together with an appropriate input, permits us to compute s .

There will always exist at least one string which is maximally incompressible. As a matter of notation, let $|s|$ denotes the length of string s . The number of programs (strings) shorter than length $|s|$ is at most $2^{|s|} - 1$ while the number of strings of size $|s|$ is $2^{|s|}$. In other words, there must be at least one string which cannot be described by a string shorter than itself for every value of $|s|$. Interestingly, as the length of a string gets large, say greater than around 1000, the vast majority of strings of that length will be maximally incompressible. These strings can be considered random in that they exhibit no pattern that can be exploited to compress them. An algorithm for generating such strings can do no better than to simply list the string. *RANDOM* (the set of all random bit strings) and *INCOMPRESSIBLE* (the set of all incompressible bit strings) are equivalent for the processes we are examining.

However, can even Demon decide for all bit strings whether it is an element of *INCOMPRESSIBLE*? Is there an effective procedure which Demon can use to establish for any given string whether it is maximally incompressible? Such a procedure would be of more than purely theoretical interest.⁸ A person whose choices are described by an algorithmically random (maximally incompressible) string of 0s (not rational) and 1s (rational) should most certainly fail a test for being rational.

To illustrate compressibility, Figure 1 shows results of simulating the choice bit strings Demon might cal-

⁸Indeed, the existence of such a procedure would imply a solution to the halting problem!

culate for agents selecting a toothpaste from 50 available brands. In order to arrive at an agent’s total order, Demon would compute preferences and choices over 1,225 distinct pairs. As above, there are then 2^{1225} or $\sim 5.8 \times 10^{368}$ distinct bit strings which could possibly result from a given agent’s choices. Generating this number of strings is well beyond the capacity of the computer we are using so our simulation generates, with replacement, pseudo-random samples of the strings. The bit string for a person whose preferences formed a total order over the 50 brands and who unerringly chose positively with regard to those preferences would be a string of 1,225 1s. The plot in 1 shows the results for different likelihoods of discrepancies from making rational choices for 1000 agents for discrepancy likelihoods ranging from 0 to 1 in .1 increments. This was done 1225 times for each of the 1000 agents being simulated at each likelihood of deviating from a rational choice.⁹

The plot shows average byte size for compressed choice bit strings at each discrepancy likelihood level. Strings were compressed using Mathematica’s built in lossless compression algorithm, COMPRESS[19]. The fences around each mean denote the maximum and minimum compressed byte sizes at each likelihood level. As expected, mean deviation likelihoods from 0 (no deviations) to .5 (random choosing) monotonically increase and beginning at .1, there is overlap between the maximum compressed size at one level and the minimum compressed size at the next level.

Byte size is an imperfect proxy for string size as the specific byte count will depend upon the compression algorithm used as well as specifics of how data are stored on a particular machine. Once compressed, a 1225 length bit string of all 1s describing each choice as rational is reduced to 64 bytes. To that must be added a constant, c , reflecting the size of the compression/decompression code. That program code can be thought of as a string of bits encoding the patterns looked for by the particular compression algorithm. Similarly, particular *ATRC* sentences can be expressed as a string of bits which, among other things, can be used to predict patterns in choice data. Those predictive

⁹These 1000 choice bit strings are an exceedingly small sample of the $\sim 5.8 \times 10^{368}$ set of all possible bit strings of length 1225. The proportion is so close to 0 that attempting to display it as an approximate real number signals a “too small to represent as a normalized machine number” warning on our machine.

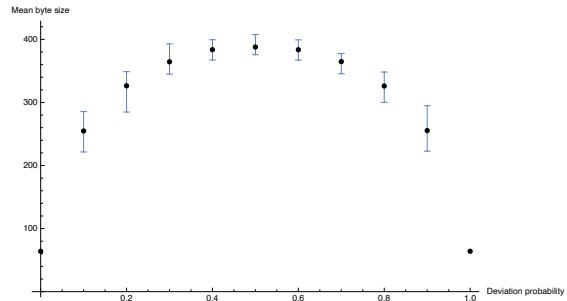


Figure 1: Mean Byte Size for 50 Alternatives as Function of n

sentences will also have a constant bit size the value of which, c , will depend upon the precise encoding used. In the example of Figure 1, the size of the compression code, c , is the same for each choice bit string so we can look at the relative compressibility of the strings without knowing the value of c .¹⁰ The importance of c will become more apparent below.

As discrepancy likelihoods go from .5 to 1 (always deviate from a rational choice), we see a mirror image of what happened between 0 and .5. Interestingly these are data Demon will have since, by assumption, Demon has perfect knowledge of both preferences and choices. As mere human observers without direct access to preference information, an agent with a deviation likelihood of 1 would look to us like a rational chooser whose preferences are revealed as reversed from what Demon knows them to be. From Demon’s perspective agents whose likelihood of deviating from a rational choice are greater than .5 would appear as *perverse* choosers in that their preferences inform their choices though in a negative way.¹¹ The irrational agents would be those choosing randomly. They are irrational in that there is no informative relation between their preferences and their choices. Knowing with certainty a random chooser’s preferences provides no advantage in predicting his choices.

Figure 1 is intended to illustrate the relation pattern between the compressibility of a choice string and the propensity of an agent to exhibit discrepancies between preferences and choices. The specific amount of com-

¹⁰The source code for COMPRESS has not been made available though it appears to be based on *Zlib*[20].

¹¹Oddly enough, reversibility of preferences does not always work as expected. [21] shows that reversibility has unexpected consequences with an *ATRC* based upon Luces’s choice axiom.

pression we can get for a given bit string will depend in part upon the specific compression algorithm used. A particular compression algorithm gives an initial upper bound estimate regarding how much a bit string can be compressed. There may well be another algorithm, perhaps yet to be discovered, that will do better. Put differently, if the *ATRC* we are using fails to see compressible patterns in the choice bit string, perhaps another, more complex, *ATRC* would reveal those patterns. All this must be understood knowing that that for long bit strings the vast majority will *not* be compressible.

There is an important sense in which Figure 1 is misleading as it displays choice bit strings generated from known probability distributions. The assumption of stochastic randomness, that is that each choice bit string at each discrepancy level are independent and identically distributed, is not compatible with the individual choice problem as we have described it. Demon calculates a choice bit string only using knowledge of an agent's preferences and choices. There is no known underlying distribution of deviations for any particular agent and no expectation that any two agents would deviate from rationality in accord with a common random process. For a given agent, Demon mechanically calculates a single choice bit string of finite length and must then decide whether that string is patterned or if it is random. Demon is in the domain of algorithmic randomness and must first decide whether the single calculated choice bit string is random and thus describes choices from an irrational chooser. Equivalently, Demon has to decide whether the bit string is compressible.

A theorem due to Chaitin [22] proves there can be no effective procedure for deciding that any given bit string is maximally incompressible. Given a formal system, T (and recall that any *ATRC* theory would be such a T), then there is a number c , depending upon T such that there is no effective procedure derivable from T that can decide whether the statement asserting " $K(s)$, the complexity of string s , is greater than c " is true. What is surprising is that we know such complex strings must exist by the simple counting argument above. Indeed, as the length of strings gets long, the vast majority will be incompressible. Thus there exist statements of the form "the complexity of string s is greater than c " that are undecidable in T . Even Demon, assuming it to be a Turing compatible entity, will not be able to decide whether every string of choices is irrational, in the sense of being complex g random and thus incompressible, under a

given *ATRC*. Importantly, even if we attempt to augment a particular *ATRC* with additional assumptions, as long as the resulting theory continues to be consistent and sound, the result will continue to hold though perhaps with a different value of c . Put differently, no matter how many axioms we add or refine, there will continue to exist choice bit strings which we can not prove to be irrational; that is, random.

In Chaitin's colorful words introducing his theorem, "...if one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from those axioms" [23, p. 942]. The theorem can be proved by first assuming that membership in *INCOMPRESSIBLE* is decidable by a Turing machine M . Since *INCOMPRESSIBLE* was produced by M its elements are denumerable and can be ordered lexicographically. We can then construct another Turing machine, M^* that takes as input a decimal number, n , and looks sequentially through elements of *INCOMPRESSIBLE* and returns s_n , the lexicographically first bit string in *INCOMPRESSIBLE*, which is of length n . The binary representation of the decimal number n will be $\approx \log_2(n)$ bits long. That is, s_n , an element of *INCOMPRESSIBLE*, is uniquely described by the bit string made up of the code for M^* followed by the binary representation of n or $s_n \equiv \{ \langle M^* \rangle \langle n \rangle \}$.

We also know that since s_n produced by M^* in an n -bit string element of *INCOMPRESSIBLE*, $K(s_n) \geq n$. The length of $s_n \equiv \{ \langle M^* \rangle \langle n \rangle \}$ which describes s_n is equal to the length of M^* which is a constant, c once M^* is fixed plus the length of the binary string encoding n , $\log_2(n)$.

Taken together, this yields: $n \leq \log_2(n) + c$. This inequality cannot be true any sufficiently large value of c . Therefore we conclude that Turing machine M cannot exist and membership in the set *INCOMPRESSIBLE* is undecidable. In turn this means *RANDOM* is also undecidable.

The impossibility of an effective procedure does not mean that it is never possible to decide whether a finite length choice bit string is random. In fact, for certain choice strings such as those containing only 1s, this will be straightforward. Also, for choice strings that are short relative to a given *ATRC*, proper determinations may be possible. The impossibility of an effective procedure means that there can be no decision algorithm derivable from *any* *ATRC* which is guaranteed to work for any choice bit string it encounters. Demon will, in

general, be unable to decide whether a choice bit string came from an irrational agent.

3. Conclusion

We argue that for a person to be described as rational does not mean that **all** her choices are rational. Mistakes, for example, are possible. Rather, rationality requires at minimum that one's choices not be random. In this paper, we show that even for a Turing compatible Demon with full knowledge of an agent's preferences and choices, it will be undecidable whether or not any given sequence of that agent's choices is random in the algorithmic complexity sense of being incompressible. This result reflects the challenges of conceptualizing rational choice while respecting computability constraints. As has been pointed out by others [3], [6], [8], [21], [24], much extant rational choice theory has been developed using the mathematics of real analysis with little concern for whether the underlying alternative sets and choice functions were computable under the Church-Turing thesis.

While this result suggests that we will not have axiomatic theory capable of deciding the rationality of all possible strings of choices, it also illustrates that considering the compressibility of these strings will provide upper bounds on their algorithmic complexity. Those bounds can improve as new theories identify patterns not derivable from previous ones. Compressibility measures, as defined within particular theories, offer computable metrics for assessing relative rationality even as the Kolmogorov complexity of a string is not computable.

Rational choice theorists generally use classical stochastic models in reasoning about uncertainty in the decision process. While often appropriate, they can be misleading when there is no estimable underlying distribution from which observed data were drawn. In such cases, algorithmic randomness, as discussed in this paper, becomes significant.

Public policy problems have been analyzed representing preferences with real valued utility functions and using real analysis to identify optimal policies. Alternatives are modeled as an uncountably infinite set of real numbers most of which are not computable. If utility functions are restricted to computable reals, many optimality arguments, which depend upon topological

fixed point theorems, break down [6], [8]. As the number of alternatives grows, axiom sets would have to be ever more complex if they are to be of use in deciding whether choice strings are non-random in situations where each choice string is *sui generis*.

It is not an adequate defense to argue that axiomatic rational choice theories merely posit that agents behave "as if" they could make these calculations and it thus is not important to know whether they actually do. This would have us focus only on predictions and not worry about the underlying mechanisms required for the theory to make those predictions. However, it is precisely those mechanisms that are significant in addressing counterfactual questions arising in debates over best decisions.

On the one hand, axiomatic theories of rational choice yield powerful descriptions of choice behavior. On the other, this power comes at the expense of axioms which can be brittle with regard to computational limits. The kind of conclusion reached in this paper is not unusual as impossibility and undecidability results have become standard fare post Gödel [25] and Turing [15]. As an increasing number of decision support models become implemented on digital computers, it is important to scrutinize underlying assumptions of computability. We suggest the challenge is to develop theories of human decision-making and rational choice which have computability built firmly into their axiomatic foundations.

4. References

- [1] D. Austen-Smith and J. S. Banks, *Positive political theory*. University of Michigan Press, 1999, vol. 2. DOI: 10.3998/mpub.105376.
- [2] H. A. Simon, "A behavioral model of rational choice," *The Quarterly Journal of Economics*, vol. 69, no. 1, p. 99, Feb. 1955, ISSN: 0033-5533. DOI: 10.2307/1884852.
- [3] P. Bossaerts, N. Yadav, and C. Murawski, "Uncertainty and computational complexity," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1766, p. 20180138, Dec. 2018, ISSN: 1471-2970. DOI: 10.1098/rstb.2018.0138.

- [4] R. Selten, “Bounded rationality,” *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, vol. 146, no. 4, pp. 649–658, 1990.
- [5] M. K. Richter, K.-C. Wong, *et al.*, “What can economists compute?” Society for Computational Economics, Tech. Rep., 1999.
- [6] M. K. Richter and K.-C. Wong, “Computable preference and utility,” *Journal of Mathematical Economics*, vol. 32, no. 3, pp. 339–354, 1999. DOI: 10 . 1016 / S0304 - 4068 (98) 00063-9.
- [7] K.-C. Wong and M. K. Richter, “Bounded rationalities and computable economies,” Discussion paper 297, Minnesota-Center for Economic Research, Tech. Rep., 1996.
- [8] K. V. Velupillai, *Computable foundations for economics*. Routledge, 2012. DOI: 10 . 4324 / 9780203002070.
- [9] R. D. Luce, *Individual choice behavior*. Dover Publications New York, 2005. DOI: 10 . 1037 / 14396-000.
- [10] L. J. Savage, *The foundations of statistics*. John Wiley & Sons, 1954. DOI: 10 . 1002 / nav . 3800010316.
- [11] H. Uzawa, “Note on preference and axioms of choice,” *Annals of the Institute of Statistical Mathematics*, vol. 8, no. 1, pp. 35–40, 1956. DOI: 10 . 1007 / bf02863564.
- [12] K. J. Arrow, “Rational choice functions and orderings,” *Economica*, vol. 26, no. 102, pp. 121–127, 1959. DOI: 10 . 2307 / 2550390.
- [13] M. K. Richter, “Revealed preference theory,” *Econometrica: Journal of the Econometric Society*, pp. 635–645, 1966. DOI: 10 . 2307 / 1909773.
- [14] S. Aaronson, *Quantum computing since Democritus*. Cambridge University Press, 2013. DOI: 10 . 1017 / CBO9780511979309.
- [15] A. M. Turing, “On computable numbers, with an application to the entscheidungsproblem,” *Proceedings of the London mathematical society*, vol. 2, no. 1, pp. 230–265, 1937. DOI: 10 . 1112 / plms / s2-42 . 1 . 230.
- [16] H. G. Rice, “Recursive real numbers,” *Proceedings of the American Mathematical Society*, vol. 5, no. 5, pp. 784–791, 1954, ISSN: 00029939, 10886826. DOI: 10 . 1090 / s0002-9939-1954-0063328-5.
- [17] P. Simon, M. de Laplace, F. Truscott, and F. Emory, “A philosophical essay on probabilities,” *The Mathematical Gazette*, vol. 2, no. 37, p. 245, 1903. DOI: 10 . 1090 / s0002-9939-1954-0063328-5.
- [18] M. Li and P. Vitányi, *An introduction to Kolmogorov complexity and its applications*, 3rd ed. Springer, 2008. DOI: 10 . 1007 / 978-0-387-49820-1_2.
- [19] W. Research, *Mathematica Edition: Version 8.01*. Champaign, Illinois: Wolfram Research, Inc., 2012.
- [20] P. Deutsch and J.-L. Gailly, “Zlib compressed data format specification version 3.3,” RFC 1950, May, Tech. Rep., 1996. DOI: 10 . 17487 / rfc1950.
- [21] S. J. Thorson and J. Stever, “Classes of models for selected axiomatic theories of choice,” *Journal of Mathematical Psychology*, vol. 11, no. 1, pp. 15–32, 1974. DOI: 10 . 1016 / 0022-2496 (74) 90009-1.
- [22] G. J. Chaitin, “On the length of programs for computing finite binary sequences,” *Journal of the ACM (JACM)*, vol. 13, no. 4, pp. 547–569, 1966. DOI: 10 . 1145 / 321356 . 321363.
- [23] ———, “Gödel’s theorem and information,” *International Journal of Theoretical Physics*, vol. 21, no. 12, pp. 941–954, Dec. 1982, ISSN: 1572-9575. DOI: 10 . 1007 / bf02084159.
- [24] K. V. Velupillai, *Models of Simon*. Routledge, 2017. DOI: 10 . 4324 / 9780203462447.
- [25] K. Gödel, “On formally undecidable propositions of Principia Mathematica and related systems i (1931),” *Monatshefte für Mathematik und Physik*, 1931. DOI: 10 . 1007 / bf01700692.