# Rational inductive agents

Caspar Oesterheld
Duke University
Durham, NC, USA
caspar.oesterheld@duke.edu

Abram Demski
Machine Intelligence Research Institute
San Francisco, CA, USA

Vincent Conitzer
Duke University
Durham, NC, USA

## ABSTRACT

The dominant theories of rational decision making assume what we will call logical omniscience. That is, they assume that when facing a decision problem, an agent can perform all relevant computations and determine the truth value of all relevant logical/mathematical claims. This assumption is unrealistic when, for example, we offer bets on remote digits of $\pi$ or Goldbach's conjecture; or when an agent faces a computationally intractable planning problem. Furthermore, the assumption of logical omniscience creates contradictions in cases where the environment can contain descriptions of the agent itself. Importantly, strategic interactions as studied in game theory are decision problems in which a rational agent is predicted by its environment (the other players). In this paper, we develop a theory of rational decision making that does not assume logical omniscience. We consider agents who repeatedly face decision problems (including ones like betting on Goldbach's conjecture or games against other agents). The main contribution of this paper is to provide a sensible theory of rationality for such agents. Roughly, we require that a rational inductive agent tests each efficiently computable hypothesis infinitely often and follows those hypotheses that keep their promises of high rewards. We then prove that agents that are rational in this sense have other desirable properties. For example, they learn to value random and pseudo-random lotteries at their expected reward. Finally, we consider strategic interactions between different agents and show that under suitable independence assumptions, rational inductive agents can converge only to playing a Nash equilibrium against each other.

**ACM Reference Format:**
Caspar Oesterheld, Abram Demski, and Vincent Conitzer. 2021. Rational inductive agents. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 20 pages.

## 1 INTRODUCTION

The dominant theories of rational decision making – in particular Bayesian theories – assume logical omniscience, i.e., that rational agents can determine the truth value of any relevant logical statement. In some types of decision problems, this prevents one from deriving any recommendation from these theories, which is unsatisfactory. For one, there are problems in which computing an optimal choice is simply computationally intractable (Section 3.1). For example, many planning problems are intractable. Second, the assumption of logical omniscience creates contradictions (resembling classic paradoxes of self reference, such as the liar's paradox) if the environment is allowed to contain references to the agent itself (Section 3.2). These issues arise most naturally when multiple rational agents interact and reason about one another.

Drawing on ideas from previous work by Garrabrant et al. [9], this paper develops a novel theory of rational inductive agents (RIAs) that does not assume logical omniscience and yields sensible recommendations in problems like the ones described above. Rather than describing how an agent should deal with an individual decision, the theory considers how an agent learns to choose on a sequence of different decision problems. We describe the setting in more detail in Section 2.

The core of our theory is a normative rationality criterion for such learning agents. Roughly, the criterion requires that a rational inductive agent test each efficiently computable hypothesis (or more generally each hypothesis in some class) infinitely often and follows those hypotheses that keep their promises of high rewards. We describe the criterion in detail in Section 4. Importantly, the criterion can be satisfied by computationally bounded agents, as we show in Section 5.

We demonstrate the appeal of our criterion by showing that it implies various desirable and general behavioral patterns. In Section 6, we show that on sequences of decision problems in which one available option guarantees a payoff of at least $l$, RIAs learn to obtain a reward of at least $l$. Thus, in particular, they avoid money extraction schemes (in the limit). In Section 7, we show that similarly on sequences of decision problems in which one available option pays off (truly or pseudo-)randomly with mean $\mu$, RIAs learn to obtain a reward of at least $\mu$. In Section 8, we show that we can construct RIAs that are also RIAs on subsequences of the sequence of decision problems faced. Finally, we consider decision problems in which one RIA plays a strategic game against another RIA. We show that under suitable independence assumptions, if RIAs converge to some strategy profile, that strategy profile must be a Nash equilibrium. Related work is discussed in Section 10. Throughout this paper, we describe the key ideas for our proofs in the main text. Detailed proofs are given in Appendix A.

## 2 SETTING

We study a very general form of what has been called a contextual multi-armed bandit problem (see Section 10 for a discussion of that literature). Let $\mathcal{T}$ be some language describing available *options*. A *decision problem* $\mathrm{DP} \in \mathrm{Fin}(\mathcal{T})$ is a finite set of options. In this paper, we will often consider specific and somewhat unusual types of decision problems as examples, in particular ones where options are terms in some mathematical logic. However, our theory applies at least as well to more traditional decision problems. For example, one could imagine that each option $a \in \mathrm{DP}$ describes a particular medical treatment (including the empirical evidence for that treatment, and so on) and that the agent has to select one of the treatments for a particular patient.

*Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.*

A *decision problem sequence* consists of a sequence of decision problems $\mathrm{DP}_1, \mathrm{DP}_2, \ldots$ along with a sequence $\bar{D}$ of functions $D_t : \mathrm{DP}_t \rightsquigarrow [0, 1]$ that at each time $t$ resolve the decision problem $\mathrm{DP}_t$ by (potentially non-deterministically) assigning a *reward* to each of the options in $\mathrm{DP}_t$. We also allow the $\mathrm{DP}_t$ and $D_t$ to depend on the history in arbitrary (including adversarial) ways. They need not be drawn from some fixed distribution or the like.

At each time $t$, an agent chooses from $\mathrm{DP}_t$, according to some *policy* $c_t : \mathrm{Fin}(\mathcal{T}) \rightsquigarrow \mathcal{T}$, i.e., a function that from any given decision problem DP selects one of the available options $c_t(\mathrm{DP}) \in \mathrm{DP}$. The agent receives feedback about its choice. We assume that at the very least the agent is informed of $D_t(c_t(\mathrm{DP}_t))$, its reward for its choice in this round, though other information may also be revealed at time $t$. The policy $c_t$ at time $t$ can depend on all information revealed up to time $t$. We focus on learning myopically optimal behavior. That is, we want our agent to learn to choose whatever gives the highest reward for the present decision problem, regardless of what consequences that has for future decision problems.

# 3 LOGICAL UNCERTAINTY AND COMPUTATIONAL CONSTRAINTS

In this paper, our goal is to develop a theory that describes how a rational agent for this problem should learn to choose. The standard theory for rational decision making under uncertainty is Bayesian decision theory (BDT) (11, 18; for overviews, see, e.g., 17, 19). The main ideas of this paper are motivated by a specific shortcoming of BDT: the assumption that the agent who is subject to BDT's recommendations is logically omniscient and in particular not limited by any computational constraints. We aim to develop a theory that can give recommendations for computationally bounded agents. In the following, we give two different kinds of examples to illustrate the role of logical omniscience in Bayesian decision theory and motivate our search for an alternative theory.

## 3.1 Mere intractability

The first problem with BDT is that in most realistic choice problems, it is hopelessly intractable to follow BDT. Full Bayesian updating or Bayes-optimal decision making are themselves only feasible if the environment is small or highly structured (18, Sections 2.5, 5.5; 7; 6). Note that even if the agent automatically had a perfectly accurate world model, then determining the optimal choice may require an agent to solve a variety of computational problems, such as vehicle routing, the traveling salesman problem, etc. In such problems, BDT simply requires an agent to find the optimal solution. However, we would like a theory of rational choice that is able to make recommendations for realistic, bounded agents who can only solve such problems approximately.

For illustration, we give two examples of decision problems in which it is especially clear what recommendation we expect from a theory of rationality for computationally bounded agents. First, consider a decision problem $\mathrm{DP} = \{a_1, a_2\}$, where the agent knows that option $a_1$ pays off 1 if the $10^{100}$-th digit of the binary representation of $\pi$ is odd and 0 otherwise. Option $a_2$ pays off 0.6 with certainty. All that Bayesian decision theory has to say about this problem is that one should calculate whether the $10^{100}$-th digit of $\pi$ is odd – if it is, choose $a_1$; otherwise choose $a_2$. Unfortunately, calculating whether the $10^{100}$-th digit of $\pi$ is odd is likely too difficult for any real-world agent.[1] Hence, Bayesian decision theory does not have any recommendations for this problem for realistic reasoners. At the same time, we have the strong normative intuition that – if digits of $\pi$ indeed cannot be predicted better than random under computational limitations – it is rational to take $a_2$. We would like our theory to make sense of that intuition.

A similar type of problem is that of betting on the truth of various mathematical conjectures. For example, should a rational agent rely on the unproven claim that various cryptographic methods are safe? As before, BDT has nothing to say to agents who do not have the compute to prove or disprove these claims. At the same time, it seems that even lacking such abilities, we have some intuitions for how such agents should behave. For example, most security-conscious people act as if modern cryptography was very likely to be safe.

## 3.2 Paradoxes of self-reference, strategic interactions and counterfactuals

A second problem with BDT and logical omniscience more generally is that the approach of choosing based on a logically omniscient belief system breaks if the values of different options depend on what the agent chooses. As an example, consider the following decision problem, which we will call the Simplified Adversarial Offer [after a slightly more complicated decision problem by 15]. Formally, for any policy $c$, let $\mathrm{SAO}_c = \{a_0, a_1\}$ be the decision problem where $a_0$ is known to pay off 0.1 with certainty, and $a_1$ is known to pay off 1 if $c(\mathrm{SAO}_c) = a_0$, and 0 otherwise. That is, the reward for $a_1$ is 1 if and only if $c$ does *not* choose $a_1$.

Now imagine that $c$ (deterministically) makes the optimal choice given a logically omniscient belief system. Then the agent knows the value of each of the options. This also means that it knows the value of $c(\mathrm{SAO}_c)$. But given this knowledge, $c$ selects a different option than what the belief system predicts. This is a contradiction. Hence, there exists no policy (that can be described in $\mathcal{T}$ and resolved by $\bar{D}$) that complies with standard BDT in this problem.

We are particularly interested in problems in which this failure mode applies. SAO is an extreme and unrealistic example, selected to be simple and illustrative. However, strategic interactions between different rational agents share the ingredients of this problem: Agent 1 is thinking about what agent 2 is choosing, thereby creating a kind of reference to agent 2 in agent 2's environment. Further, it may be in agent 2's interest to prove wrong whatever agent 1 believes about agent 2. For a closely related discussion of issues of bounded rationality and the foundations of game theory, see Binmore [4] and references therein.

# 4 A RATIONALITY CRITERION

In this section, we propose our novel rationality requirement, which is the key contribution of this paper. In short, our approach is as follows: Agents have to not only choose, but also estimate the reward they will receive. As part of our rationality criterion we require

[1] Remote digits of $\pi$ are the canonical example in the literature on logical uncertainty. To the knowledge of the authors it is not known whether the $n$-th digit of $\pi$ can be guessed better than random in less than $O(n)$ time. For a general, statistical discussion of the randomness of digits of $\pi$, see Marsaglia [13].

that these estimates are not systematically above what the agent actually obtains. Further, we consider rationality relative to some set of hypotheses, which in turn make recommendations and claim that some utility is achieved when following the recommendation. To satisfy computational constraints, we can restrict the set of hypotheses to only include efficiently computable ones. Roughly, our rationality criterion then states that if a hypothesis claims strictly higher utility than the agent infinitely often, then the agent must test this hypothesis infinitely often. Testing requires taking the option recommended by the hypothesis in question. To reject a hypothesis, these tests must indicate that the hypothesis consistently over-promises.

## 4.1 Preliminary definitions

A *decision market* $\bar{\alpha}$ is a sequence of functions $\alpha_t\colon \mathrm{Fin}(\mathcal{T}) \to \mathcal{T} \times [0,1]\colon \mathrm{DP}_t \mapsto (\alpha^c(\mathrm{DP}_t), \alpha^e(\mathrm{DP}_t))$. Here, $\alpha^c$ is a policy (with $\alpha^c(\mathrm{DP}_t) \in \mathrm{DP}_t$ as before). We will call the values of $\alpha^e$ estimates. For example, we might like a decision market for which $\alpha_t(\mathrm{SAO}_{\alpha_t}) = (a_0, 0.1)$, which means that the agent chooses $a_0$ (which pays 0.1 with certainty) and estimates that it will receive a reward of 0.1. Again, we imagine that the sequence $\bar{\alpha}$ is constructed by facing $\mathrm{DP}_t$ at each time $t$ and receiving feedback from $D_t$.

A *bidder* $\bar{b}$ has the same type signature as a decision market, but we will use it in a different role than decision markets, namely that of a hypothesis that a decision market has to test. When taking about bidders, we will often refer to the values of $b_t^e$ as promises.

Our rationality criterion will be relative to a particular set of bidders $\mathbb{B}$. In principle, $\mathbb{B}$ could be any set of bidders, e.g. all computable functions, all three-layer neural nets, etc. Generally, $\mathbb{B}$ should contain any bidder (i.e., any hypothesis about how the agent should act) that the decision market agent is willing to consider, similar to the support of the prior in Bayesian theories of learning. Following Garrabrant et al. [9], we will often let $\mathbb{B}$ be the set of functions computable in $O(g(t))$ time, where $g$ is a function with $g(t) \to \infty$ as $t \to \infty$. We will call these functions *efficiently computable (e.c.)*.

Besides a set of bidders, our rationality criterion will be relative to some countable set of subsequences of the sequence of decision problems $\bar{D}$. The full reason for this will only become clear later. We discuss this in detail in Appendix B. Roughly, we may like an agent to be rational not only on $\bar{D}$, but also on any (potentially sparse) efficiently decidable subsequence of $\bar{D}$. This is particularly relevant if we imagine that a large fraction of the training problems are quite different from real-world problems.

We will denote this set of subsequences of $\bar{D}$ by $\mathbb{S}$. We assume that $\mathbb{S}$ is countable. Each subsequence $S \in \mathbb{S}$ is simply a subset $S \subseteq \mathbb{N}$ of indices $t$ at which $\mathrm{DP}_t$ is part of the subsequence. As with bidders, we will often imagine the subsequences $S \in \mathbb{S}$ are decidable in $O(g(t))$. We will call such subsequences *efficiently decidable (e.d.)*. Two important special cases of $\mathbb{S}$ are the one where $\mathbb{S}$ contains each e.d. subsequence and the case where $\mathbb{S} = \{\mathbb{N}\}$, i.e., where we do not account separately for different subsequences at all.

## 4.2 Low absolute loss

We now describe the first part of our rationality requirement, which is that the estimates should not be systematically above what the agent actually obtains. The criterion itself is straightforward, but its significance will only become clear in the context of the low relative loss criterion of the next section.

**Definition 1.** *We call*

$$\mathcal{L}_T(\bar{\alpha}, \bar{D}, S) \coloneqq \sum_{t \in S_{\leq T}} \alpha_t^e(\mathrm{DP}_t) - D_t(\alpha_t^c(\mathrm{DP}_t)) \tag{1}$$

*the cumulative absolute loss of a decision market $\bar{\alpha}$ on the subsequence $S$ of $\bar{D}$.*

Now the low absolute loss criterion states that the average per-round absolute loss is at most 0 in the limit:

**Definition 2.** *Let $S$ be an infinite subsequence of $\bar{D}$. We say $\bar{\alpha}$ for $\bar{D}$ has low absolute loss on $S$ if $\mathcal{L}_T(\bar{\alpha}, \bar{D}, S)/|S_{\leq T}| \leq 0$ as $T \to \infty$.*

In other words, for all $\epsilon > 0$, there should be a time $t$ such that for all $T > t$, $\mathcal{L}_T(\bar{\alpha}, \bar{D}, S)/|S_{\leq T}| \leq \epsilon$. Note that the per-round absolute loss of rational inductive agents as defined below will usually but need not always converge to 0; it can be negative in the limit, see Appendix B.

## 4.3 Low relative loss

We now come to our second requirement, which specifies how the agent $\bar{\alpha}$ relates to the bidders in $\mathbb{B}$. This second requirement is more complicated in its general form, so we will describe it more piece-wise. We start with the following definition.

**Definition 3.** *We say that $\bar{b}$ outbids $\bar{\alpha}$ or that $\bar{\alpha}$ rejects $\bar{b}$ at time $t$ if $b_t^e(\mathrm{DP}_t) > \alpha_t^e(\mathrm{DP}_t)$.*

Intuitively, we might distinguish two kinds of bidders: First, there are bidders that promise higher rewards than $\bar{\alpha}^e$ in only finitely many rounds. For example, this will be the case for hypotheses that $\bar{\alpha}$ is taking into account when choosing and estimating. Also, this could include bidders who recommend an inferior option with an accurate estimate, e.g., bidders who recommend "0.05" and promise 0.05 in {"0.05", "0.1"}. For all of these bidders, we do not require anything from $\bar{\alpha}$. In particular, $\bar{\alpha}$ need not test these bidders.

Second, some bidders do infinitely often promise strictly higher rewards than $\bar{\alpha}^e$. For these cases, we will require of our rational inductive agents that they have some reason to distrust these bidders. To be able to provide such a reason, $\bar{\alpha}$ needs to test these bidders infinitely often. Testing a bidder requires following that bidder's recommendation.

**Definition 4.** *We call an infinite set $M$ a test set of $\bar{\alpha}$ for $\bar{b}$ if for all $t \in N$, $\alpha_t^c(\mathrm{DP}_t) = b_t^c(\mathrm{DP}_t)$.*

For $\bar{\alpha}$ to infinitely often reject $\bar{b}$, these tests must then show that $\bar{b}$ is not to be trusted (in those rounds in which they promise a reward that exceeds $\bar{\alpha}^e$). That is, on these tests, the rewards must be significantly lower than what the bidder promises. We thus introduce another key concept.

**Definition 5.** *Let $\bar{b}$ be a bidder and $N \subseteq \mathbb{N}$ be a test set. We call*

$$l_T(\bar{\alpha}, \bar{D}, N, \bar{b}) \coloneqq \sum_{t \in N_{\leq T}} D_t(b_t^c(\mathrm{DP}_t)) - b_t^e(\mathrm{DP}_t) \tag{2}$$

*$b$'s empirical record on $N$.*

We now have all the pieces together to state our definition. For illustration, we first give our definition for the special case of $\mathbb{S} = \{\mathbb{N}\}$, i.e., where the decision market does not account separately by subsequence: Let $\bar{\alpha}$ be a decision market, $\bar{b}$ be a bidder and let $B$ be the set of times $t$ at which $\bar{\alpha}$ rejects $\bar{b}$. We say that $\bar{\alpha}$ *has low relative loss (without subsequence accounting) to* $\bar{b}$ on a test set $M$ if either $B$ is finite or $l_T(\bar{\alpha}, \bar{D}, M, \bar{b}) \to -\infty$ as $T \to \infty$ within $B$.

Relative to the general definition below, the only idea missing from the definition in the previous paragraph is subsequence accounting. We will allow an agent $\bar{\alpha}$ to maintain separate empirical records for different subsequences and to reject a bidder $\bar{b}$ at time $t$ if $t \in S_j$ for some subsequence $S_j \in \mathbb{S}$ and $\bar{b}$ has a poor empirical record on $M \cap S_j$, where $M$ is a test set of the agent for $\bar{b}$. In the following definition, the function $\rho$ will point for each round $t$ in which $\bar{\alpha}$ rejects $\bar{b}$ to the subsequence containing $t$ on which $\bar{b}$ has a poor empirical record, thus justifying the rejection of $\bar{b}$ at time $t$. Conversely, we can use $\rho^{-1}$ to get for a given subsequence $S_j$ the set of times $t$ at which $S_j$ is used as the reason for rejection.

Subsequence accounting comes with a new challenge. The above definition of low relative loss puts no constraints on rejecting any given bidder finitely many times. But now imagine that we account for an infinite set of subsequences $\mathbb{S}$. A naive approach might similarly allow for each subsequence $S_j$, that $\bar{\alpha}$ can finitely many times give $S_j$ as a reason for rejecting $\bar{b}$, without any restriction on the empirical record of $\bar{b}$ on $M \cap S_j$. But then $\bar{\alpha}$ could reject any given bidder $\bar{b}$ all the time, simply by pointing at a different subsequence $S_j \in \mathbb{S}$ at each instance.

To address this problem, we will introduce a function $f$, which for each subsequence $S_j \in \mathbb{S}$ gives a (minimum) tolerance for $S_j$. A bidder can then only be rejected based on its record on $M \cap S_j$ if that record is worse than $-f(j)$. We must then let $f$ grow fast enough. In particular, we will require $f \in \omega(\log)$, for reasons that will become clear in Section 7.1. We still require that for each individual subsequence the tolerance goes to $\infty$. That is, if a subsequence $S_j$ is named as the reason for rejecting $b$ infinitely often, then the record of $b$ on $M \cap S_j$ must become arbitrarily bad.

**Definition 6.** *Fix a particular enumeration $S_1, S_2, \ldots$ of $\mathbb{S}$ and a function $f(\cdot) \in \omega(\log)$. Let $B$ be the set of times $t$ at which $b$ strictly outbids $\alpha$. We say that $\alpha$ has low relative loss to $b$ with subsequence accounting for $\mathbb{S}$ if either $B$ is finite or if along with $\alpha$ we can compute an infinite test set $M$ of $\bar{\alpha}$ for $\bar{b}$ with the following property. There must be a function $\rho\colon B \to \mathbb{N}$ that at each time $T \in B$ specifies the index of a subsequence $\rho(T)$ that $T$ is an element of, i.e., $T \in S_{\rho(t)}$, such that for each $T \in B$ it is*

$$l_T(\bar{\alpha}, \bar{D}, M \cap S_{\rho(T)}, \bar{b}) < -f(\rho(T)). \tag{3}$$

*Further, if $\rho^{-1}(j)$ is infinite then the empirical record of $\bar{b}$ on $M \cap S_j$, i.e., $l_T(\bar{\alpha}, \bar{D}, M \cap S, \bar{b})$, must go to $-\infty$ as $T \to \infty$ among $T \in \rho^{-1}(j)$.*

If $\bar{\alpha}$ does not have low relative loss to $\bar{b}$, we will also say that $\bar{b}$ *exploits* $\bar{\alpha}$.

Note that we allow $f$ to assume negative values. At first sight, this may be counterintuitive because it means ineq. 3 is trivially satisfied if $b_i$ has not been tested on that subsequence at all. However, for each bidder $\bar{b}_i$, using an $f_i$ with negative values only allows this unjustified rejection of $\bar{b}_i$ finitely many times.

### 4.4 The rational inductive agent criterion

**Definition 7.** *We say that a decision market $\bar{\alpha}$ with test sets $M_1, M_2, \ldots$ is a rational inductive agent (RIA) for $\bar{D}, \mathbb{B} = \{b_1, b_2, \ldots\}, \mathbb{S} = \{S_1, S_2, \ldots\}, f_1, f_2, \ldots$ if it has low relative loss on test set $M_i$ to bidder $b_i \in \mathbb{B}$ and tolerance function $f_i$ for all $i$ and low absolute loss on any infinite subsequence $S \in \mathbb{S}$ of $\bar{D}$.*

In the following, whenever $\bar{\alpha}$ is a RIA, we will imagine that the test sets are given as a part of $\bar{\alpha}$. For example, if we say that $\bar{\alpha}$ is computable in $O(h(t))$, then we will take this to mean that $\bar{\alpha}$ together with a list at time $t$ of tested bidders can be computed in that time.

## 5 COMPUTING RATIONAL INDUCTIVE AGENTS

As described in Section 3, the goal of this paper is to formulate a rationality requirement that is not self-contradictory and that can be satisfied by computationally bounded agent. Therefore, we must show that one can actually construct RIAs for given $\mathbb{S}$ and $\mathbb{B}$ and that under some assumptions about $\mathbb{S}$ and $\mathbb{B}$, such RIAs are computable within some (asymptotic) bounds.

Theorem 1. *Let $\mathbb{B}$ and $\mathbb{S}$ be enumerable and $\bar{D}$ be some decision process. (Let $g \in \Omega(\log)$.) Then if $\mathbb{B}$ and $\mathbb{S}$ can be computably enumerated and consist only of ($O(g(t))$-)computable bidders and subsequences, then there exists a RIA for $\mathbb{B}$, $\mathbb{S}$, $\bar{D}$ that is computable (in $O(g(t)q(t))$, for arbitrarily slow-growing, $O(g(t))$-computable $q$ with $q(t) \to \infty$).*

It can similarly be shown that, for example, a RIA relative to the class $P$ of bidders computable in polynomial time can be computed in arbitrarily close to polynomial time, i.e. in $O(x^{q(t)})$ for arbitrarily slow-growing $q$ with $q(t) \to \infty$. The next result shows that the RIAs given by Theorem 1 are asymptotically optimal in terms of runtime.

Theorem 2. *Let $\alpha$ be a RIA for $\bar{D}, \mathbb{B}$. Assume that there are infinitely many $t$ such that $|\mathrm{DP}_t| \geq 2$ and $\alpha_t^e(\mathrm{DP}_t) < 1$. If $\mathbb{B}$ is the set of ($O(g(t))$-)computable bidders, then $\alpha$ is not computable (in $O(g(t))$).*

This is shown by a simple diagonalization argument. If a RIA $\alpha$ were computable (in $O(g(t))$), then consider the bidder who in rounds in which $|\mathrm{DP}_t| \geq 2$ and $\alpha_t^e(\mathrm{DP}_t) < 1$, bids 1 and recommends an option other than $\alpha_t^c(\mathrm{DP}_t)$; and bids 0 otherwise. This bidder strictly outbids $\alpha$ infinitely often, is computable (in $O(g(t))$) but can never be tested.

## 6 EASY OPTIONS

Throughout most of the rest of this paper, we will show that RIAs satisfy many desiderata that one might have for rational decision makers. In this section, we start with a simple result which shows that if one of the options can be efficiently shown to have a value of at least $L_t$, then a RIA will come to expect to obtain at least $L_t$ on that subsequence.

Theorem 3. *Let $\bar{D}$ be a decision problem sequence and $\bar{\alpha}$ be a RIA for $\bar{D}$ and the set of e.c. bidders. Let $S$ be an e.d. subsequence of $\bar{D}$ and $\bar{a}$ be a sequence of terms in $\mathcal{T}$ s.t. for $t \in S$ it is $a_t \in \mathrm{DP}_t$ and $\alpha_t^c(\mathrm{DP}_t) = a_t \implies D_t(a_t) \geq L_t$ for some e.c. sequence $\bar{L}$. We require*

*also that the $a_t$ are efficiently identifiable from the set* $\mathrm{DP}_t$*. Then for all but finitely many $t \in S$ it is $\alpha_t^e(\mathrm{DP}_t) \geq L_t$. As a consequence, if $\alpha$ has low absolute loss on $S$ (e.g., because $\alpha$ accounts for $S$), then in the limit as $T \to \infty$ it is*

$$\frac{1}{|S_{\leq T}|} \sum_{t \in S_{\leq T}} D_t(\alpha_t^c(\mathrm{DP}_t)) \geq \frac{1}{|S_{\leq T}|} \sum_{t \in S_{\leq T}} L_t. \tag{4}$$

The last claim of the theorem must be conditional on low loss on $S$ for the reasons given in Appendix B.

The proof idea is simple. Consider the bidder who estimates $L_t$ and recommends $a_t$ if $t \in S$ and bids 0 otherwise. This bidder never overestimates. Hence, to ensure low relative loss to this bidder, $\alpha$ can be strictly outbid by this bidder only finitely many times.

Theorem 3 implies that when the value of all options is e.c., then (assuming appropriate subsequence accounting) a RIA must choose the best available option. For example, when the choice is between "0.05" and "0.1", a RIA has to choose "0.1" with frequency 1.

We can also interpret Theorem 3 as providing an immunity to money extraction schemes, which is one of the most widely discussed rationality conditions. If a RIA can leave with a certain payoff of $L_t$, it will on average leave with at least $L_t$. It will not converge to giving up the $L_t$ for some other in-the-limit inferior option. For example, in the Simplified Adversarial Offer of Section 3.2, a RIA must walk away with at least 0.1, which in turn means that it must choose option $a_0$ = "0.1" with frequency 1. As Oesterheld and Conitzer [15] show, a different normative theory of rationality, called causal decision theory, can be used as a money pump.

Another corollary of Theorem 3 is that RIAs must learn and use empirical facts that can be efficiently deduced from what is revealed by $\bar{D}$. For example, imagine that in one round, $\bar{D}$ reveals that the minimum of the populations of Hamburg and Amsterdam is 0.8 million. Then in later rounds, this information can be used to efficiently compute lower bounds on other options. For example, the option that pays off the maximum of the populations of Hamburg and Detroit in millions can be deduced to be at least 0.8. If such decision problems occur infinitely often, RIAs must converge to exploiting such inferences.

# 7 LOTTERIES

## 7.1 True randomness

The following result shows, roughly, that in the limit RIAs are vNM-rational whenever being vNM-rational is possible given the computational restrictions. That is, when choosing between different lotteries and the expected utilities of these lotteries can be computed efficiently, RIAs converge to choosing the lottery with the highest expected utility. When other, non-lottery options are available, RIAs must converge to performing at least as well as the best lottery option.

Theorem 4. *Let $\bar{D}$ be a decision problem sequence and $\alpha$ be a RIA for $\bar{D}$ relative to some $f_1, f_2, \ldots$ with $f_i \in \omega(\log)$ for all $i$. Let $R \subseteq \mathbb{N}$ be an e.d. subsequence. Let $\bar{a}$ be a sequence of terms in $\mathcal{T}$ s.t. for $t \in R$, $a_t \in \mathrm{DP}_t$ and the values $D_t(a_t)$ are drawn independently from some distributions with e.c. means $\bar{\mu}$. We further require that the $a_t$ are efficiently identifiable from* $\mathrm{DP}_t$*. Then almost surely in the limit as $t \to \infty$, it is $\alpha_t^e(\mathrm{DP}_t) \geq \mu_t$. As a consequence, if $\alpha$ has low absolute loss on $R$, then in the limit as $T \to \infty$ it is*

$$\frac{1}{|R_{\leq T}|} \sum_{t \in R_{\leq T}} D_t(\alpha^c(\mathrm{DP}_t)) \geq \frac{1}{|R_{\leq T}|} \sum_{t \in S_{\leq T}} \mu_t. \tag{5}$$

We here give an intuition, which also explains in more detail than before where the $\omega(\log)$ comes from. The basic idea is similar to the idea behind the proof of Theorem 3: For any $\epsilon > 0$, we consider the bidder who recommends $a_t$ and estimates $\mu_t - \epsilon$. It is clear that on any individual subsequence, this bidder will almost surely have a positive record in the limit. The hard part is showing that there will almost surely be only finitely many subsequences on which the bidder at some point has a negative record exceeding $f$. Here we need the requirement that $f_i(j)$ grows at some minimum speed in $j$ for any bidder $i$. Note that if, for example, the $D_t(a_t)$ are unbiased coin-flips, $R = \mathbb{N}$ and $\alpha$ tests the present bidder in every round, then as $T \to \infty$, there will almost surely be some subsequence of past decision problems on which the bidder has amassed an overall overestimate of about $T/4$. (Namely, the subsequence that happens to consist of the roughly half ($T/2$) of the decision problems on which 0 was obtained and the bidder bid almost 1/2.) We need to ensure that the probability that one of these "ex post adversarial" sequences has an $O(T)$ bound goes to zero sufficiently quickly. Hence, of the $2^T$ subsequences of the first $T$ rounds, almost all should have an order $\omega(T)$ tolerance. This is ensured exactly by requiring that $f_i \in \omega(\log)$ for each bidder $i$.

## 7.2 Pseudo-randomness

Theorem 4 only tells us something about *true* random variables. But a key goal of our theory is to also be able to assign expected rewards to *pseudo*-random sequences, i.e., sequences that are deterministic and potentially even computable, but impossible to predict better than a random sequence under computational constraints.

The main obstacle in deriving such a version of Theorem 4 lies in defining pseudo-randomness itself. Many notions of pseudo-randomness have been discussed in the literature (e.g. 10, Ch. 8; 5). However, these capture somewhat different concepts than what is relevant for us. The following is tailored to what is needed for our theorem.

**Definition 8.** *We call a sequence $\bar{a}$ as resolved by $\bar{D}$ pseudo-random with means $\bar{\mu}$ if for all efficiently enumerable sequences $S_1, S_2, \ldots$ of subsequences that are efficiently decidable given $\alpha$'s computations, all functions $f$ from some family of functions $\mathcal{F} \subseteq \omega(\log)$ and all $\epsilon > 0$, it is only for finitely many times $T$ the case that there is a $j$ for which*

$$\sum_{t \in S_{j,\leq T}} \mu_t - \epsilon - D_t(a_t) > f(j). \tag{6}$$

Essentially this definition states that with the computational power of $\bar{\alpha}$, it is impossible to consistently pick out subsequences on which the actual values of $a_t$ are lower than its mean [cf. 9, Definition 4.1.1].

Theorem 5. *Let $\bar{D}$ be a decision problem sequence and $\alpha$ be a RIA for $\bar{D}$. Let $N \subseteq \mathbb{N}$ be an e.d. subsequence. Let $\bar{a}$ be pseudo-randomly evaluated by $\bar{D}$ with e.c. means $\bar{\mu}$ and with a family of functions $\mathcal{F}$ that includes for each bidder $b_j$ the functions $f(j, \cdot)$ from the RIAs low relative loss criterion. We further require that the $a_t$ are efficiently identifiable from* $\mathrm{DP}_t$*. Then in the limit as $t \to \infty$, it is $\alpha_t^e(\mathrm{DP}_t) \geq \mu_t$.*

This follows directly from the proof of Theorem 4.

## 8 RIAS ON SUBSEQUENCES

Imagine that we train a RIA $\bar{\alpha}$ on a sequence of decision problems $\bar{D}$ before that agent faces a real-world problem. At the time of training, we might not yet know what type of real-world problem the agent will face. We may thus be inclined to include a variety of training problems. For example, we might train the agent both for interactions with other sophisticated and potentially adversarial agents as well as for single-agent decision scenarios. Then when the real-world decision problems do not involve other agents, we would hope that when viewing the sequence of decision problems without other agents, $\bar{\alpha}$ is still a RIA. RIAs indeed have this property, if we account for the respective subsequence. (Again, see Appendix B on why subsequence accounting is (and should be) necessary.)

Theorem 6. *Let $\bar{\alpha}$ be a RIA for $\bar{D}$, $\mathbb{B}$, $\mathbb{S}$. We require that $\mathbb{B}$ and $\mathbb{S}$ satisfy the following closure properties relative to each other: 1) For each $b \in \mathbb{B}, S \in \mathbb{S}$ there is a $b' \in B$ s.t. $b'_t(\mathrm{DP}_t) = b_t(\mathrm{DP}_t)$ whenever $t \in S$, and $b^e_t(\mathrm{DP}_t) = 0$, otherwise. 2) For all $S, S' \in \mathbb{S}$, $S \cap S' \in \mathbb{S}$. Now, let $N \in \mathbb{S}$. Then $\bar{\alpha}_{|N}$ is a RIA for $\bar{D}_{|N}$ which accounts for subsequences $\{S \in \mathbb{S} \mid S \subseteq N\}$.*

Here, $\bar{\alpha}_{|N}$ is $\alpha$ restricted to elements of $N$.

As a corollary of Theorem 6, RIAs that account for all e.d. subsequences exhibit self-similarity through the lense of the RIA criterion.

**Corollary 7.** *Let $\bar{\alpha}$ be a RIA for $\bar{D}$ which accounts for all e.d. subsequences. Let $N$ be an e.d. subsequence. Then $\bar{\alpha}_{|N}$ is a RIA for $\bar{D}_{|N}$ that accounts for all e.d. subsequences.*

## 9 RIAS AS A FOUNDATION FOR GAME THEORY

We start by briefly giving definitions of the relevant game-theoretic concepts. For a thorough introduction to game theory, refer to Osborne [16] or any other textbook on the topic. A *(two-player) game* consists of two finite sets of *(pure) strategies* $A_1, A_2$, one set for each player, and two payoff functions $u_1, u_2 \colon A_1 \times A_2 \to [0, 1]$. A *(pure) strategy profile* is a pair $(a_1, a_2) \in A_1 \times A_2$. We call $(a_1, a_2)$ a *(pure) Nash equilibrium* if for $i = 1, 2$, $a_i \in \arg\max_{a'_i \in A_i} u_i(a'_i, a_{-i})$. We call the Nash equilibrium strict if both argmaxes are singletons.

Now we imagine that we have two RIAs $\bar{\alpha}, \bar{\beta}$. For simplicity, assume that they do not account separately for subsequences and only play the particular game (rather than interspersing other decision problems). One possible way of expressing the game in decision problem sequences is as follows. For each $a_1 \in A_1$, $\mathrm{DP}^{\alpha}_t$ contains an option that pays $u_1(a_1, a_2)$, where $a_2$ is the action corresponding to the option selected by $\beta_t$ in $\mathrm{DP}^{\beta}_t$. And $\mathrm{DP}^{\beta}_t$ is defined analogously. Abusing notation a little, we use $a_i \in A_i$ to represent the available options in $\mathrm{DP}^{\alpha/\beta}_t$. For instance, we write $\alpha^c_t(\mathrm{DP}^{\alpha}) = a_1$ to denote that $\alpha_t$ chooses the option from $\mathrm{DP}^{\alpha}_t$ that corresponds to $a_1 \in A_1$.

How two RIAs $\alpha, \beta$ play against each other, depends on what exactly these RIAs look like. In particular, if $\alpha, \beta$ always choose the same, then by Theorem 3 they will converge to cooperate in the Prisoner's Dilemma, thus deviating from Nash equilibrium play. However, we believe that when two RIAs do not happen to be correlated in this way, they will generally only be able to converge to playing a Nash equilibrium (if they converge at all).

We will here prove this under an assumption about non-correlation between $\bar{\alpha}$ and $\bar{\beta}$. We start by defining what it means for one player's test set to be uncorrelated in the relevant sense with the other player's choices. We say that a set $M \subseteq \mathbb{N}$ is *weakly uncorrelated* with $\bar{\beta}$ if whenever $\beta^c_t(\mathrm{DP}_t) = a_2$ with frequency 1 on $t \in \mathbb{N}$ for some $a_2 \in A_2$, $\beta^c_t(\mathrm{DP}_t) = a_2$ is also true with frequency 1 on $t \in M$.

Next, we define the kind of bidder that prevents convergence to non-equilibria, if tested in an uncorrelated way. Let $a_2 \in A_2$ and $a^*_1 \in \arg\max_{a_1} u_1(a_1, a_2)$ be a best response to $a_2$. Also let $\mu = \max\left(\{u_1(a_1, a_2) \mid a_1 \in A_1\} - \{u_1(a^*_1, a_2)\}\right)$ be the utility for Player 1 of playing a second-best response to $a_2$. Then we call $b$ a safe $a_2 \to a^*_1$ bidder if there is an $\epsilon > 0$ s.t. if $\beta^c_t(\mathrm{DP}_t) = a_2$ with frequency 1, then $b^c_t(\mathrm{DP}_t) = a^*_1$ and $b^e(\mathrm{DP}_t) \in \left[\mu + \epsilon, u_1(a^*_1, a_2) - \epsilon\right]$ with frequency 1 and otherwise $b^e_t(\mathrm{DP}_t) = 0$.

**Assumption 1.** *Assume that for each pair of $a_2 \in A_2$ and a best response $a^*_1$ to $a_2$, there is a safe $a_2 \to a^*_1$ bidder who either outbids $\bar{\alpha}$ only finitely many times or whose test set in $\bar{\alpha}$ is weakly uncorrelated with $\bar{\beta}$.*

Under this assumption, we can show that convergence is only possible to Nash equilibria.

Theorem 8. *Let $\bar{\alpha}, \bar{\beta}$ be RIAs for the decision problem sequences $\bar{D}^{\alpha}, \bar{D}^{\beta}$, respectively, and no subsequence accounting (i.e., $\mathbb{S}_{\alpha/\beta} = \{\mathbb{N}\}$). If $\alpha^c_t(\mathrm{DP}_t), \beta^c_t(\mathrm{DP}_t)$ converge to choosing with frequency 1, the options corresponding to $a_1 \in A_1, a_2 \in A_2$, then under Assumption 1 for both $\alpha$ relative to $\beta$ and $\beta$ relative to $\alpha$, $(a_1, a_2)$ is a Nash equilibrium of the underlying game.*

It is not immediately obvious whether Assumption 1 will naturally (e.g., when the two RIAs are designed independently without attempts to coordinate) be satisfied or not. However, we conjecture that convergence to non-equilibria is rare and generally requires fine tuning the RIAs to each other in some way.

Finally, we show that for every strict Nash equilibrium, there is a pair of RIAs that converge to that Nash equilibrium.

Theorem 9. *For each game $(A_1, A_2, u_1, u_2)$ and strict Nash equilibrium $(a_1, a_2) \in A_1 \times A_2$ there is a pair of randomizing decision markets $\bar{\alpha}, \bar{\beta}$ that are RIAs with probability 1 relative to any (countable) set of bidders $\mathbb{B}$ (and without subsequence accounting) and that converge to playing $(a_1, a_2)$ with probability 1.*

## 10 RELATED WORK

*Decision theory of Newcomb-like problems.* Problems in which the environment explicitly predicts the agent have been discussed as Newcomb-like problems by (philosophical) decision theorists [2, 14].

Most of this literature has focused on discussing relatively simple cases (similar to the Simplified Adversarial Offer). In these cases, RIAs generally side with what has been called evidential decision theory. For example, by Theorem 3, RIAs learn to one-box in Newcomb's problem. Of course, RIAs differ structurally from how a decision theorist would usually conceive of an evidential decision theory-based agent. E.g., RIAs are not based on expected utility

maximization (though they implement it when feasible, see Section 7.1). We also note that the decision theory literature has, to our knowledge, not produced any formal account of how to assign the required conditional probabilities in Newcomb-like problems.

**Contextual stochastic multi-armed bandits** As noted in Section 2, our problem is a contextual multi-armed bandit problem as considered in statistical learning theory. However, papers in this literature generally avoid the possibility of an environment that can refer to the agent (as in the Adversarial Offer or strategic interactions). For example, Yang and Zhu [21, Assumption A in Section 5] and Agarwal et al. [1, Assumption 1 in Section 2] assume that the agent's models can converge to being accurate. These assumptions allow a much simpler rationality requirement, namely some kind of convergence to optimal behavior (cf. Section 6).

**Adversarial multi-armed bandits with expert advice** Another closely related literature is that on multi-armed bandit problems with expert advice (3, Section 7; 12, Chapter 18). This literature generally allows adversarial problems. Like this paper, it addresses this problem by making the optimality goal relative to some set of hypotheses. However, its optimality criterion is quite different from ours: they require regret minimization and in particular that regret converges to 0, a condition sometimes called Hannan-consistency. As the Simplified Adversarial Offer shows, Hannan-consistency is not achievable in our setting. However, it does become achievable if we assume that the agent has access to a source of random noise that is independent from $\bar{D}$ [see, e.g, the Exp4 algorithm of 3, Section 7].

We find it implausible to *require* rational agents to randomize to minimize regret; most importantly, regret minimization can require minimizing the rewards one actually obtains, see Appendix C. At the same time, we conjecture that learners with low regret relative to a set of bidders $\mathbb{B}$ satisfy a version of the RIA criterion, see Appendix D for a preliminary result.

**Garrabrant inductors** As noted earlier, the present approach to dealing with computational constraints is inspired by the work of Garrabrant et al. [9], who address the problem of assigning probabilities under computational constraints. As an alternative to the present theory of RIAs, one could also try to develop a theory of decision making under logical uncertainty by maximizing expected utility, using the Garrabrant inductor's probability distributions. Unfortunately, this approach fails for reasons related to the challenge of making counterfactual claims, as pointed out by Garrabrant [8]. As in the case of Hannan consistency, we can address this problem using randomization over actions. However, like Garrabrant (ibid.), we do not find it satisfactory to *require* randomization (cf. again Appendix C). We conjecture that, like regret minimizers, Garrabrant inductors with (pseudo-)randomization could be used to construct RIAs.

## 11 CONCLUSION

We developed a RIA theory as a theory of bounded rationality. We gave results that show the normative appeal of RIAs. Furthermore, we demonstrated the theory's utility by using it to justify Nash equilibrium play. At the same time, the ideas presented lead to various further research questions, some of which we have noted above. We here give three more that we find particularly interesting. Can we modify the RIA requirement so that it implies coherence properties à la Garrabrant et al. [9]? Do the frequencies with which RIAs play the given pure strategies of a game converge to mixed Nash and correlated equilibria? Can RIA theory be used to build better real-world systems?

## REFERENCES


[1] Alekh Agarwal, Miroslav Dudík, Satyen Kale, John Langford, and Robert E. Schapire. 2012. Contextual Bandit Learning with Predictable Rewards. In *Appearing in Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*. Proceedings of Machine Learning Research, Vol. 22. La Palma, Canary Islands, 19–26. http://proceedings.mlr.press/v22/agarwal12/agarwal12.pdf

[2] Arif Ahmed. 2014. *Evidence, Decision and Causality.* Cambridge University Press.

[3] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Shapire. 2001. The non-stochastic multi-armed bandit problem. https://cseweb.ucsd.edu/~yfreund/papers/bandits.pdf

[4] Ken Binmore. 1987. Modeling Rational Players: Part I. *Economics & Philosophy* 3, 2 (10 1987), 179–214.

[5] Gregory J. Chaitin. 2001. *Exploring Randomness.* Springer.

[6] Krishnendu Chatterjee, Martin Chmelík, and Mathieu Tracol. 2016. What is decidable about partially observable Markov decision processes with $\omega$-regular objectives. *J. Comput. System Sci.* 82, 5 (8 2016), 878–911.

[7] Gregory F. Cooper. 1990. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence* 42, 2-3 (3 1990), 393–405. https://doi.org/10.1016/0004-3702(90)90060-D

[8] Scott Garrabrant. 2017. Two Major Obstacles for Logical Inductor Decision Theory. https://www.alignmentforum.org/posts/5bd75cc58225bf06703753d4/two-major-obstacles-for-logical-inductor-decision-theory

[9] Scott Garrabrant, Tsvi Benson-Tilsen, Andrew Critch, Nate Soares, and Jessica Taylor. 2016. Logical Induction. https://intelligence.org/files/LogicalInduction.pdf A short version is available at https://intelligence.org/files/LogicalInductionAbridged.pdf; another short version was published in TARK '17, see https://arxiv.org/abs/1707.08747.

[10] Oded Goldreich. 2008. *Computational Complexity: A Conceptual Perspective.* Cambridge University Press.

[11] Richard C. Jeffrey. 1965. *The Logic of Decision.* McGraw-Hill, New York.

[12] Tor Lattimore and Csaba Szepesvari. 2017. Bandit Algorithms. (2017). https://tor-lattimore.com/downloads/book/book.pdf

[13] George Marsaglia. 2005. On the Randomness of Pi and Other Decimal Expansions. *InterStat* (10 2005). http://interstat.statjournals.net/YEAR/2005/articles/0510005.pdf

[14] Robert Nozick. 1969. Newcomb's Problem and Two Principles of Choice. In *Essays in Honor of Carl G. Hempel*, Nicholas Rescher et al. (Ed.). Springer, 114–146. http://faculty.arts.ubc.ca/rjohns/nozick_newcomb.pdf

[15] Caspar Oesterheld and Vincent Conitzer. 2021. Extracting Money from Causal Decision Theorists. *The Philosophical Quarterly* (2021). https://doi.org/10.1093/pq/pqaa086

[16] Martin J. Osborne. 2004. *An Introduction to Game Theory.* Oxford University Press.

[17] Martin Peterson. 2009. *An Introduction to Decision Theory.* Cambridge University Press.

[18] Leonard J. Savage. 1954. *The Foundations of Statistics.* John Wiley and Sons, New York.

[19] Katie Steele and H. Orri Stefánsson. 2016. Decision Theory. In *The Stanford Encyclopedia of Philosophy* (winter 2016 ed.), Edward N. Zalta (Ed.). https://plato.stanford.edu/archives/win2016/entries/decision-theory/

[20] Paul Weirich. 2016. Causal Decision Theory. In *The Stanford Encyclopedia of Philosophy* (spring 2016 ed.).

[21] Yuhong Yang and Dan Zhu. 2002. Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *The Annals of Statistics* 30, 1 (2002), 100–121.

# A PROOFS

## A.1 An easy lemma about test sets

We start with a simple lemma which we will use to simplify a few of our proofs.

**Lemma 10.** *Let $\bar{b}$ be a bidder and $N \subseteq \mathbb{N}$ s.t. $t \in N$ implies $b_t^e(\mathrm{DP}_t) = 0$. Then if $M$ is a test set that ensures low relative loss of $\bar{\alpha}$ to $\bar{b}$, so is $M - N$.*

Proof. For all $T$, it is

$$
\begin{aligned}
& l_T(\bar{\alpha}, \bar{D}, M, \bar{b}) && (7)\\
= & \sum_{t \in M_{\leq T}} D_t(b_t^c(\mathrm{DP}_t)) - b_t^e(\mathrm{DP}_t) && (8)\\
= & \sum_{t \in M_{\leq T} - N} D_t(b_t^c(\mathrm{DP}_t)) - b_t^e(\mathrm{DP}_t) + \sum_{t \in M_{\leq T} \cap N} D_t(b_t^c(\mathrm{DP}_t)) - b_t^e(\mathrm{DP}_t) && (9)\\
= & \sum_{t \in M_{\leq T} - N} D_t(b_t^c(\mathrm{DP}_t)) - b_t^e(\mathrm{DP}_t) + \sum_{t \in M_{\leq T} \cap N} D_t(b_t^c(\mathrm{DP}_t)) && (10)\\
\geq & \sum_{t \in M_{\leq T} - N} D_t(b_t^c(\mathrm{DP}_t)) - b_t^e(\mathrm{DP}_t) && (11)\\
= & \; l_t(\bar{\alpha}, \bar{D}, M - N, \bar{b}). && (12)
\end{aligned}
$$

Thus, if $l_T(\bar{\alpha}, \bar{D}, M, \bar{b}) \to -\infty$ as $T \to -\infty$, it must also be $l_T(\bar{\alpha}, \bar{D}, M - N, \bar{b}) \to -\infty$ as $T \to -\infty$. □

## A.2 Proof of Theorem 1

Theorem 1. *Let $\mathbb{B}$ and $\mathbb{S}$ be enumerable and $\bar{D}$ be some decision process. (Let $g \in \Omega(\log)$.) Then if $\mathbb{B}$ and $\mathbb{S}$ can be computably enumerated and consist only of ($O(g(t))$-)computable bidders and subsequences, then there exists a RIA for $\mathbb{B}$, $\mathbb{S}$, $\bar{D}$ that is computable (in $O(g(t)q(t))$, for arbitrarily slow-growing, $O(g(t))$-computable $q$ with $q(t) \to \infty$).*

Proof. Our proof is divided into four parts. First, we give the generic construction for a RIA (1). Then we show that this is indeed a RIA by proving that it satisfies the low absolute loss condition (2), as well as the low relative loss condition (3). Finally, we show that under the assumptions stated in the theorem, this RIA is computable in the claimed time complexity (4).

1. The construction We first need to select for each bidder $b_i$ our function $f_i$, which specifies what we will call the initial *wealth* of the bidder in the algorithm and which will map onto the $f$ from the low relative loss criterion (Definition 6). The following criteria need to be satisfied:

- For each bidder $b_i \in \mathbb{B}$, it must be the case that $f(i, \cdot) \in \omega(\log)$ as specified in the low relative loss condition.
- For each subsequence $S_j \in \mathbb{B}$, the series

$$\sum_{i=1}^{\infty} f_i(j) + 1 \tag{13}$$

  must converge absolutely to some finite value.

An example of an admissible function $f$ is $f_i(j) = ji^{-2} - 1$.

Second, we need an *allowance function* $A : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \to \mathbb{R}_{\geq 0}$, which for each subsequence $S_j$ and each subsequence-specific time/count $n$, specifies a positive amount $A(n, i, j)$ given to bidder $b_i$'s subsequence-specific wealth when the $n$-th (lowest) element of $S_j$ occurs in $\bar{D}$. The allowance function must satisfy the following requirements:

- Each bidder must get infinite overall allowance on each infinite subsequence, i.e.,

$$\sum_{n=1}^{\infty} A(n, i, j) = \infty \tag{14}$$

  for all bidders $b_i$ and all subsequences $S_j$.
- The overall average allowance distributed per round $n$ on any particular subsequence must go to zero, i.e.,

$$\sum_{n=1}^{N} \frac{1}{N} \sum_{i=1}^{\infty} A(n, i, j) \underset{N \to \infty}{\to} 0 \tag{15}$$

for all $S_j \in \mathbb{S}$. In particular, the allowance distributed in any particular round of any particular subsequence must be finite.

An example of such a function is $A(n, i, j) = n^{-1} i^{-2}$.

We can finally give the algorithm itself. Initialize the wealth variables $w_0(i, j) \leftarrow f_i(j) + 1$ for each bidder $b_i \in \mathbb{B}$ and subsequence $S_j \in \mathbb{S}$.

At time $t$, we run a (first-price sealed-bid[2]) auction for the present decision problem among all bidders. That is, we determine a winning bidder

$$i_t^* \in \arg\max_{i \in \mathbb{N}} \min(b_{i,t}^e(\mathrm{DP}_t), \min_{j:\, t \in S_j} w_t(i, j)) \tag{16}$$

with arbitrary tie-breaking. Intuitively, each bidder $b_i$ bids $b_{i,t}^e(\mathrm{DP}_t)$, except that it is constrained by the wealths $w_t(i, j)$ it has on any subsequence $S_j$ that $t$ belongs to. Let $e_t^*$ be the maximum (wealth-bounded) bid itself. We then define our decision market at time $t$ to do $\alpha_t(\mathrm{DP}_t) := (b_{i_t^*,t}^c(\mathrm{DP}_t), e_t^*)$.

We update the wealth variables as follows. For all $j$ with $t \notin S_j$ and all $i$, the wealth simply stays the same, i.e., $w_{t+1}(i, j) \leftarrow w_t(i, j)$. For all $j$ with $t \in S_j$ and all bidders $i \neq i_t^*$, we merely give allowance, i.e.

$$w_{t+1}(i, j) \leftarrow w_t(i, j) + A(|S_j \cap \{1, ..., t\}|, i, j). \tag{17}$$

For the winning bidder $i_t^*$ and all $j$ with $t \in S_j$, we update wealth according to

$$w_{t+1}(i_t^*, j) \leftarrow w_t(i_t^*, j) + A(|S_j \cap \{1, ..., t\}|, i_t^*, j) + D_t(b_{i_t^*,t}^c(\mathrm{DP}_t)) - e_t^*. \tag{18}$$

That is, the winning bidder receives the allowance and the reward obtained after following its recommendation ($D_t(b_{i_t^*,t}^c(\mathrm{DP}_t))$), but pays its bid ($e_t^*$).

2. Low absolute loss We will show that the absolute loss on any sequence $S_j \in \mathbb{S}$ is bounded by the sum of the allowance and initial wealth for that subsequence.

For each $j, T$, let $B_{j,T}^+$ be the set of bidders whose wealth $w_t(i, j)$ is non-negative for at least one time $t \in \{0, ..., T\}$. Note that all highest bidders in rounds 1...., $T$ are in $B_{j,T}^+$ for all $j$. We can then write the overall wealth of the bidders in $B_{j,T}^+$ at time $T$ for subsequence $S_j$ as

$$\begin{aligned} \sum_{i \in B_{j,T}^+} w_T(i, j) = &\sum_{i \in B_{j,T}^+} f_i(j) + 1 \\ &+ \sum_{i \in B_{j,T}^+} \sum_{n=1}^{|S_{j,\leq T}|} A(n, i, j) \\ &+ \sum_{t \in S_{j,\leq T}} D_t(\alpha_t^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t). \end{aligned} \tag{19}$$

That is, the overall wealth for subsequence $S_j$ is the initial wealth on that subsequence plus the allowance distributed on that subsequence plus the money earned by the highest bidders on that subsequence.

Now notice that by the construction above, if a wealth variable $w_t(i, j)$ is non-negative once, nit remains non-negative for all future $t$. Thus, for all $i \in B_{j,T}^+$, $w_T(i, j) \geq 0$. Second, the last term is the negated absolute loss of $\bar{\alpha}$ on $S_j$. Thus, re-arranging these terms and dividing by $|S_{j,\leq T}|$ gives us the following upper bound on the per-round absolute loss:

$$\frac{1}{|S_{j,\leq T}|} \mathcal{L}_T(\alpha, \bar{D}, S_j) \tag{20}$$

$$= \frac{1}{|S_{j,\leq T}|} \left( \left( \sum_{i \in B_{j,T}^+} f_i(j) + 1 \right) + \sum_{i \in B_{j,T}^+} \sum_{n=1}^{|S_{j,\leq T}|} A(n, i, j) - \sum_{i \in B_{j,T}^+} w_T(i, j) \right) \tag{21}$$

$$\leq \frac{1}{|S_{j,\leq T}|} \left( \sum_{i \in B_{j,T}^+} f_i(j) + 1 \right) + \frac{1}{|S_{j,\leq T}|} \sum_{i \in B_{j,T}^+} \sum_{n=1}^{|S_{j,\leq T}|} A(n, i, j) \tag{22}$$

We now argue that as $T \to \infty$, if $S_j$ is infinite, then this upper bound converges to 0. First, notice that we have required $\sum_{i=1}^{\infty} f_i(j) + 1$ to converge absolutely to a finite value. This implies (by some well-known ideas

[2]This format is mainly chosen for its simplicity. We could just as well use a second-price (or third-price, etc.) auction. We could use even different formats to get somewhat different RIA-like properties. For instance, with combinatorial auctions, one could achieve cross-decision optimization.

from real analysis) that the sums $\sum_{i \in B^+_{j,T}} f_i(j) + 1$ are bounded. Thus, as $|S_{j,\leq T}| \to \infty$, the first summand of the upper bound converges to 0. Second, because $A$ is non-negative,

$$\frac{1}{|S_{j,\leq T}|} \sum_{i \in B^+_{j,T}} \sum_{n=1}^{|S_{j,\leq T}|} A(n,i,j) \leq \sum_{i=1}^{\infty} \frac{1}{|S_{j,\leq T}|} \sum_{n=1}^{|S_{j,\leq T}|} A(n,i,j). \tag{23}$$

which goes to zero as $|S_{j,\leq T}| \to \infty$ by our requirement on the function $A$ (line 15).

3. Low relative loss Given a bidder $b_i$ who strictly outbids $\bar{\alpha}$ infinitely often, we use as a test $M_i$, the set of times $t$ at which $b_i$ is the winning bidder (i.e., $i = i^*_t$). We have to show that $M_i$ is infinite, is a valid test set (as per Definition 4), and that it satisfies the justified rejection requirement in the low relative loss criterion (Definition 6).

A) We show that $M_i$ is infinite. That is, we need to show that infinitely often $b_i$ is the highest bidder in the auction that computes $\bar{\alpha}$. Assume for contradiction that $M_i$ is finite. We will show that at some point $b_i$'s bidding in the construction of $\bar{\alpha}$ will not be constrained anymore by $b$'s wealth. We will then find a contradiction with the assumption that $b_i$ strictly outbids $\alpha$ infinitely often.

So let $T = 1 + \max(M_i)$ be such that from time $T$ on, $\bar{b}_i$ is never the winning bidder in $\bar{\alpha}$. First notice that because $f_i \in \omega(\log)$, for all but finitely many $j$, $f_i(j) \geq T$. Since in the first $T$ steps, wealth can decrease by at most 1 per round, it is for all these $j$,

$$w_T(i,j) \geq f_i(j) + 1 - T \geq 1. \tag{24}$$

Further, since $b_i$'s wealth cannot decrease if $b_i$ is not the highest bidder, it is for all $t \geq T$, $w_t(i,j) \geq 1$.

Now consider the finite set $F \subseteq \mathbb{N}$ of sequence indices $j$ for which $f_i(j) < T$. For each $j \in F$, we distinguish two cases. If $S_j$ is finite then the wealth constraints for $S_j$ become irrelevant from $T_j \coloneqq 1 + \max S_j$ onward. If $S_j$ is infinite, then consider that for $T' > T$, it is

$$w_{T'}(i,j) = w_T(i,j) + \sum_{t \in S_j : T < t \leq T'} A(|S_{j,\leq t}|, i, j). \tag{25}$$

That is, from time $T$ to any time $T'$, bidder $i$'s wealth only changes by $b_i$ receiving allowance, because $i$ is (by assumption) not the winning bidder $i^*_t$ in any round $t \geq T$. Because $S_j$ is infinite and we required $\sum_{n=1}^{\infty} A(n,i,j) = \infty$, we can select for each subsequence $j$ considered in this case a time $T_j \geq T$ such that $w_{T_j}(i,j) \geq 1$. Note that again it is also for all $t > T_j$ the case that $w_t(i,j) \geq 1$.

We now see that if

$$t \geq \max(T, \max_{j \in F} T_j), \tag{26}$$

none of the wealth constraints is actually restrictive. That is, for all such $t$ it is

$$\min(b^e_{i,t}(\mathrm{DP}_t), \min_{j:\, t \in S_j} w_t(i,j)) = b^e_{i,t}(\mathrm{DP}_t). \tag{27}$$

But it is infinitely often $b^e_{i,t}(\mathrm{DP}_t) > \alpha^e_t(\mathrm{DP}_t)$. This contradicts the fact that by construction, $\alpha_t(\mathrm{DP}_t)$ is equal to the highest wealth-restricted bidder.

B) The fact that $M_i$ is a valid test set follows immediately from the construction – $\alpha$ always chooses the recommendation of the highest bidder.

C) We come to the justification part of the low relative loss condition. Let $B_i$ be the set of rounds in which $\bar{b}_i$ strictly outbids $\bar{\alpha}$. We construct $\rho_i \colon B_i \to \mathbb{N}$ as follows. By construction of $\bar{\alpha}$, there is at each time $T \in B_i$ a subsequence $j$ such that $T \in S_j$ and $w_T(i,j) < b^e_{i,t}(\mathrm{DP}_T)$. After all, if there was no such $j$, then $b_{i,t}$ is not wealth-constrained in the auction construction of $\bar{\alpha}$, in which case it cannot be that $T \in B_i$. Let $\rho_i(T)$ return any (e.g., the lowest) such $j$.

Now consider both sides of $w_T(i,j) < b^e_{i,t}(\mathrm{DP}_T)$. It is $b^e_{i,t}(\mathrm{DP}_T) \leq 1$. Also,

$$w_T(i,j) = f_i(j) + 1 + \sum_{n=1}^{|S_{j,\leq T}|} A(n,i,j) + \sum_{t \in M_i \cap S_j : t < T} D_t(b^c_{i,t}(\mathrm{DP}_t)) - b^e_{i,t}(\mathrm{DP}_t). \tag{28}$$

Hence, from the fact that $w_T(i, \rho_i(T)) < b^e_{i,t}(\mathrm{DP}_T)$ for all $T \in B_i$, it follows that for all $T \in B_i$, it is

$$\sum_{t \in M_i \cap S_{\rho_i(T)} : t < T} b^e_{i,t}(\mathrm{DP}_t) - D_t(b^c_{i,t}(\mathrm{DP}_t)) > f(i, \rho_i(T)) \tag{29}$$

as required. Further, if $\rho^{-1}(j)$ is infinite, then for each $T \in \rho^{-1}(j)$ it is

$$\sum_{t \in M_i \cap S_j : t < T} b^e_{i,t}(\mathrm{DP}_t) - D_t(b^c_{i,t}(\mathrm{DP}_t)) > \sum_{n=1}^{|S_{j,\leq T}|} A(n,i,j), \tag{30}$$

which goes to infinity as $T \to \infty$, as required.

4. Computability and computational complexity It is left to show that if $\mathbb{B}$ and $\mathbb{S}$ can be computably enumerated and consist only of $(O(g(t))$-)computable bidders and subsequences, then we can implement the above-described RIA for $\mathbb{B}$, $\mathbb{S}$, $\bar{D}$ in an algorithm (that runs in $O(g(t)q(t))$, for arbitrarily slow-growing, $O(g(t))$-computable $q$ with $q(t) \to \infty$).

The main challenge is that the construction as described above performs at any time $t$, operations for all (potentially infinitely many) bidders and all (potentially infinitely many) subsequences. The crucial idea is that for appropriate choices of $A$ and $f_1, f_2, ...$, we only need to keep track of a finite set of bidders and subsequences, when calculating $\bar{\alpha}$ in the first $T$ time steps. For simplicity assume that $S_1 = \mathbb{N}$. We can set $f_i(1) = -1$ for all $i$. Thus, each bidder starts with an initial wealth of 0 on $\mathbb{N}$. Then a bidder $i$ can only become relevant at the first time $t$ at which $A(t,i,1) > 0$. At any time $t$, we call such bidders *active*. Before that time, we do not need to compute $\bar{b}_i$ and do not need to update its wealth. By choosing a function $A$ s.t. $A(t,\cdot,1) > 0$ has finite support at each time $t$, we can keep the set of active bidders finite at any given time. We have thus shown that it is enough to keep track at any given time of only a finite number of bidders.

Next we show that it is similarly enough to keep track of only a finite set of subsequences. For each bidder $b_i$, the initial wealth $f_i(j) + 1$ goes to infinity as $j$ goes to infinity. Hence, for each bidder $i$ all but finitely many of the subsequence wealth variables will be guaranteed to be greater than $b^e_{i,t}$ and therefore irrelevant computing $\alpha_t$. Again, we call the subsequences that are relevant in this sense *active* subsequences. For example, if $f_i(j) = ji^{-2} - 1$, then at time $t$, we only need to keep track of the wealth variables for subsequences with indices $1, ..., i^2t$ because on all wealth variables for subsequences $S_j$ with index $j \geq i^2(t+1)$, the wealth will be at least $f_i(j) + 1 - t \geq 1$, since a bidder loses at most 1 unit of wealth per time step and therefore incurs an overall loss of wealth of at most $t$ in the first $t$ time steps.

At any time, we therefore only need to keep track of a finite number of wealth variables, only need to compute the recommendations and bids of a finite set of bidders, only need to decide membership in a finite set of subsequences and only need to compute a minimum of a finite set in line 16. Note that if a subsequence $j$ becomes active at time $t$, then for all active bidders $i$, we have to retroactively compute $w_t(i,j)$. We similarly have to retroactively compute $w_t(i,j)$ for all active $j$, once a bidder $b_i$ is activated.

Computability is therefore proven. We proceed to show the claim about computational complexity. To do so, we first introduce a slightly more sophisticated scheme for keeping track of wealth variables. As noted above, one simple approach is to when a bidder/subsequence is activated at time $T$, calculate its wealth variables at time $T$. However, this produces a lot of calculation at once: at time $T$ we have to test subsequence membership for all times below $T$. For sufficiently slow-growing (e.g., linear) $g$, we would have to calculate $\Theta(Tg(T))$, which wouldn't satisfy the bound of the theorem. For this proof, we therefore use a slightly more complicated calculation scheme: For notational simplicity, let all subsequences and bidders be activated in even time steps. Then if a subsequence or bidder is activated at time $2T$, we start catching up on the wealth variable updates for that subsequence or bidder at time $T$, two time steps at a time. That is, for $n = 0, 1, ...T-1$, we calculate at time $T + n$ the wealth variable updates for times $2n$ and $2n + 1$.

We now come to analyze the time complexity with that scheme. At any time $t$, let $C_{\max}(t)$ be the largest constant by which the computational complexity of active subsequences and bidders at time $t$ are bounded relative to $g(t)$. Further, let $h_s(t)$ be the number of active subsequences at time $t$, as determined by $f$ via the argument given above, and $h_b(t)$ be the set of active bidders. Then the computational cost from simulating all active bidders and subsequences at time $t$ is at most

$$2h_s(2t)C_{\max}(2t)g(t) + h_b(t)g(t). \tag{31}$$

All of $h_s(2t)$, $C_{\max}(2t)$ and $h_b(t)$ must go to $\infty$ as $t \to \infty$. However, this can happen arbitrarily slowly, up to the limits of fast $(O(g(t)))$ computation. In particular, in the case of $h_s(2t)$, we can achieve this by letting $f_i(j)$ grow very quickly in $j$. Hence, if we let $q(t) = 2h_s(2t)C_{\max}(2t) + h_b(t)$, we can let $q$ grow arbitrarily slowly (again, up to the limits of fast computation).

Finally, we have to verify that all other calculations can be done in $O(q(t)g(t))$: To determine the winning bidder given everyone's bids, we have to calculate the maximum of $h_b(t) \in O(q(t))$ numbers, each of which is the minimum of $h_s(t)+1 \in O(q(t))$ numbers. Hence, determining the winning bidder is done in $O(q(t)^2)$ and

therefore for slow-enough $q$ in $O(q(t)g(t))$. We also need to conduct the wealth variable updates themselves, which accounts for $O(h_s(2t)h_b(2t))$ additions and allowance calculations. Again, this is in $O(g(t)q(t))$. And so on. □

## A.3 Proof of Theorem 3

Theorem 3. *Let $\bar{D}$ be a decision problem sequence and $\bar{\alpha}$ be a RIA for $\bar{D}$ and the set of e.c. bidders. Let $S$ be an e.d. subsequence of $\bar{D}$ and $\bar{a}$ be a sequence of terms in $\mathcal{T}$ s.t. for $t \in S$ it is $a_t \in \mathrm{DP}_t$ and $\alpha_t^c(\mathrm{DP}_t) = a_t \implies D_t(a_t) \geq L_t$ for some e.c. sequence $\bar{L}$. We require also that the $a_t$ are efficiently identifiable from the set $\mathrm{DP}_t$. Then for all but finitely many $t \in S$ it is $\alpha_t^e(\mathrm{DP}_t) \geq L_t$. As a consequence, if $\alpha$ has low absolute loss on $S$ (e.g., because $\alpha$ accounts for $S$), then in the limit as $T \to \infty$ it is*

$$\frac{1}{|S_{\leq T}|} \sum_{t \in S_{\leq T}} D_t(\alpha_t^c(\mathrm{DP}_t)) \geq \frac{1}{|S_{\leq T}|} \sum_{t \in S_{\leq T}} L_t. \tag{4}$$

Proof. We prove the claim by proving a contrapositive. In particular, we assume that $\alpha_t^e(\mathrm{DP}_t) < L_t$ for infinitely many $t \in S$ and will then show that $\bar{\alpha}$ is not a RIA (using the other assumptions of the theorem). Consider bidder $\bar{b}_i$ such that $b_{i,t}(\mathrm{DP}_t) = (a_t, L_t)$ for $t \in S$ and $b_{i,t}^e(\mathrm{DP}_t) = 0$ otherwise. Because $S$ is e.d., $\bar{L}$ is e.c. and the $\bar{a}$ are efficiently identifiable, $\bar{b}$ is e.c. We now show that $\bar{b}_i$ exploits $\bar{\alpha}$, which shows that $\bar{\alpha}$ is not a RIA. By assumption, $\bar{b}_i$ strictly outbids $\bar{\alpha}$ infinitely often. It is left to show that there is no $M_i$ as specified in the low relative loss criterion, i.e. no $M_i$ on which $\bar{b}_i$ consistently underperforms.
By Lemma 10, we can WLOG restrict attention to $M_i$ such that $M_i \subseteq S$. Hence, if $t \in M_i$, it must be $\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t) = a_t$ and therefore $D_t(a_t) \geq L_t$. It follows that for each subsequence $S_j \subseteq \mathbb{N}$ and all $T$ it is

$$l_T(\bar{\alpha}, \bar{D}, M \cap S_j, \bar{b}_i) = \sum_{t \in M_i \cap S_j : t < T} \underbrace{b_{i,t}^e(\mathrm{DP}_t)}_{=L_t} - \underbrace{D_t(b_{i,t}^c(\mathrm{DP}_t))}_{\geq L_t} \leq 0. \tag{32}$$

It follows that the required $\rho_i$ cannot be constructed: Since $f_i(j) \to \infty$ as $j \to \infty$, $\rho_i(S)$ must be finite. But then $\rho_i^{-1}(j)$ would have to be infinite for at least one $j$. And this cannot be the case, because the partial sums have an upper bound and therefore don't go to $+\infty$. □

## A.4 Proof of Theorem 4

To prove Theorem 4, we use the following lemma, which in similar form is known as the Borel-Cantelli lemma.

**Lemma 11** (Borel-Cantelli)**.** *Let $X_1, X_2 \ldots$ be a sequence of potentially dependent binary random variables. If*

$$\sum_{i=1}^{\infty} P(X_i) \tag{33}$$

*converges to some finite value, then*

$$\sum_{i=1}^{\infty} X_i \tag{34}$$

*almost surely converges to a finite value.*

Proof. For any $K \in \mathbb{N}$ it is

$$P\left(\sum_{i=1}^{\infty} X_i = \infty\right) \leq P\left(\bigvee_{i=K}^{\infty} X_i\right) \leq \sum_{i=K}^{\infty} P(X_i). \tag{35}$$

The last step is due to the union bound a.k.a. Boole's inequality. But by assumption

$$\sum_{i=K}^{\infty} P(X_i) \to 0 \tag{36}$$

as $K \to \infty$. So it must be

$$P\left(\sum_{i=1}^{\infty} X_i = \infty\right) = 0. \tag{37}$$

□

By the way, it is not enough to assume instead of the convergence of $\sum_{i=1}^{\infty} P(X_i)$ that $P(X_i)$ goes to 0. For example, it could be that for all $k$, exactly one of $X_{2^k}, X_{2^k+1}, ..., X_{2^{k+1}-1}$ occurs and each of these occurs with the same probability $2^{-k}$. Then it is certainly (and therefore with positive probability) $\sum_i X_i = \infty$ but $P(X_i) \to 0$.
We can now prove Theorem 4, which we first restate as usual for convenience.

THEOREM 4. *Let $\bar{D}$ be a decision problem sequence and $\alpha$ be a RIA for $\bar{D}$ relative to some $f_1, f_2, ...$ with $f_i \in \omega(\log)$ for all $i$. Let $R \subseteq \mathbb{N}$ be an e.d. subsequence. Let $\bar{a}$ be a sequence of terms in $\mathcal{T}$ s.t. for $t \in R$, $a_t \in \mathrm{DP}_t$ and the values $D_t(a_t)$ are drawn independently from some distributions with e.c. means $\bar{\mu}$. We further require that the $a_t$ are efficiently identifiable from $\mathrm{DP}_t$. Then almost surely in the limit as $t \to \infty$, it is $\alpha_t^e(\mathrm{DP}_t) \geq \mu_t$. As a consequence, if $\alpha$ has low absolute loss on $R$, then in the limit as $T \to \infty$ it is*

$$\frac{1}{|R_{\leq T}|} \sum_{t \in R_{\leq T}} D_t(\alpha^c(\mathrm{DP}_t)) \geq \frac{1}{|R_{\leq T}|} \sum_{t \in S_{\leq T}} \mu_t. \tag{5}$$

PROOF. Consider for each $\epsilon > 0$ the bidder $\bar{b}_i$ who bids only on $R$, bids $\mu_t - \epsilon$ and recommends $a_t$. We assume for contradiction that this bidder outbids $\bar{\alpha}$ infinitely often. We will show that regardless of what test set $M_i$ is used, $\bar{b}_i$ almost surely would violate the justification part of the low relative loss requirement.
First, the law of large numbers implies directly that for each subsequence $S_j$,

$$l_T(\bar{\alpha}, \bar{D}, S_j \cap M_i, \bar{b}_i) = \sum_{t \in S_j \cap M_i : t < T} D_t(b_{i,t}^c(\mathrm{DP}_t)) - b_{i,t}^e(\mathrm{DP}_t) \tag{38}$$

converges with probability 1 to infinity as $T \to \infty$ if $S_j \cap M_i$ is infinite. Thus, no subsequence can be used as a reason for rejection infinitely often. If the low relative loss criterion is to be satisfied, there must therefore be infinitely many subsequences $S_j$ and times $T$ at which inequality

$$\sum_{t \in M_i \cap S_j : t < T} b_{i,t}^e(\mathrm{DP}_t) - D_t(b_{i,t}^c(\mathrm{DP}_t)) > f_i(j) \tag{39}$$

is satisfied.
For each $S_j \in \mathbb{S}$ let $X_j$ denote the binary random variable that is 1 (true) if and only if the present bidder $\bar{b}_i$ will at any time have a loss of at least $f_i(j)$ on $M_i \cap S_j$. We need to show that

$$\sum_{j=1}^{\infty} X_j \tag{40}$$

almost surely converges to some finite value. By Lemma 11, it is enough to show that

$$\sum_{j=1}^{\infty} P\left(X_j\right) \tag{41}$$

converges to a finite value.
We first derive an upper bound on $P\left(X_j\right)$, i.e., the probability that the loss to $\bar{b}_i$ is at any point at least $f_i(j)$. For any $n$, let $X_{j,n}$ be the random variable denoting that inequality 39 is satisfied for that $T$ being the $n$-th element of $S_j \cap M_i$. If $S_j \cap M_i$ has less than $n$ rounds, define $X_{i,n} = 0$ with probability 1. By the union bound (a.k.a. Boole's inequality),

$$P\left(X_j\right) \leq \sum_{n=1}^{\infty} P(X_{j,n}). \tag{42}$$

Now $X_{j,n}$ is just the probability that the sample mean of $n$ independent random variables is smaller than the mean/expected value of these random variables by $\epsilon + f_i(j)/n$. These latter probabilities can be bound by Hoeffding's inequality. In particular, it is

$$P(X_{j,n}) \underset{\text{Hoeffding}}{\leq} e^{-2n(\epsilon + f_i(j)/n)^2} \tag{43}$$

$$= e^{-2n\epsilon^2 - 2f_i(j)^2/n - 4\epsilon f_i(j)} \tag{44}$$

$$\leq e^{-2n\epsilon^2 - 4\epsilon f_i(j)}. \tag{45}$$

Note that if $n > |M \cap S_j|$, this bound is trivially satisfied.
Hence,

$$P\left(X_j\right) \leq \sum_{n=1}^{\infty} P(X_{j,n}) \leq e^{-4\epsilon f_i(j)} \sum_{n=1}^{\infty} e^{-2n\epsilon^2} = e^{-4\epsilon f_i(j)} \frac{1}{e^{2\epsilon^2} - 1}. \tag{46}$$

The last step applies the formula for the limit of a geometric series.
So,

$$\sum_{j=1}^{\infty} P\left(X_j\right) \leq \frac{1}{e^{2\epsilon^2}-1} \sum_{j=1}^{\infty} e^{-4\epsilon f_i(j)}. \tag{47}$$

This converges if $\sum_{j=1}^{\infty} e^{-4\epsilon f_i(j)}$ converges. And this is the case for all $\epsilon$ if $f_i \in \omega(\log)$. To see this, notice that

$$e^{-4\epsilon f_i(j)} = \frac{1}{\left(e^{f_i(j)}\right)^{4\epsilon}}. \tag{48}$$

If $f_i \in \omega(\log)$, then $f_i(j)$ as a function of $j$ eventually outgrows any $k\log(j) = \log(j^k)$. In particular, $f_i(j)$ eventually outgrows $\log(j^{1/(2\epsilon)})$, so at some point it starts to be the case that

$$\left(e^{f_i(j)}\right)^{4\epsilon} \geq \left(j^{\frac{1}{2\epsilon}}\right)^{4\epsilon} = j^2. \tag{49}$$

Since the series $\sum_{j=1}^{\infty} j^{-2}$ converges, the series $\sum_{j=1}^{\infty} e^{-4\epsilon f(j)}$ also converges whenever $f_i \in \omega(\log)$. □

## A.5 Proof of Theorem 6

Theorem 6. *Let $\bar{\alpha}$ be a RIA for $\bar{D}, \mathbb{B}, \mathbb{S}$. We require that $\mathbb{B}$ and $\mathbb{S}$ satisfy the following closure properties relative to each other: 1) For each $b \in \mathbb{B}, S \in \mathbb{S}$ there is a $b' \in B$ s.t. $b'_t(\mathrm{DP}_t) = b_t(\mathrm{DP}_t)$ whenever $t \in S$, and $b^e_t(\mathrm{DP}_t) = 0$, otherwise. 2) For all $S, S' \in \mathbb{S}$, $S \cap S' \in \mathbb{S}$. Now, let $N \in \mathbb{S}$. Then $\bar{\alpha}_{|N}$ is a RIA for $\bar{D}_{|N}$ which accounts for subsequences $\{S \in \mathbb{S} \mid S \subseteq N\}$.*

Proof. Low absolute loss follows directly from the fact that for $S \subseteq N$,

$$\mathcal{L}_T(\alpha_{|N}, \bar{D}_{|N}, S) = \mathcal{L}_T(\alpha, \bar{D}, S). \tag{50}$$

It is left to show that the low *relative* loss condition is satisfied for each e.c. bidder $\bar{b}_i$ on $\bar{D}_{|N}$. To prove that, consider a bidder $\bar{b}' \in \mathbb{B}$ (which exists by the first closure property) that is defined as

$$b'_t(\mathrm{DP}_t) = \begin{cases} b_t(\mathrm{DP}_t), & \text{if } t \in N \\ (*, 0), & \text{otherwise} \end{cases}, \tag{51}$$

where $*$ just means that $b^c_t(\mathrm{DP}_t)$ can simply be any option from $\mathrm{DP}_t$. Because $\bar{b}' \in \mathbb{B}$, $\bar{\alpha}$ has low relative loss to $\bar{b}'$ with some test set $M$, which by Lemma 10 is WLOG a subset of $N$, and tolerance function $f\colon \mathbb{N} \to \mathbb{R}$. Now, notice that for all $S \in \mathbb{S}$,

$$l_T(\bar{\alpha}_{|N}, \bar{D}_{|N}, M \cap (S \cap N), \bar{b}) = l_T(\bar{\alpha}, \bar{D}, M \cap S, \bar{b}'). \tag{52}$$

Further, $\bar{\alpha}_{|N}$ rejects $\bar{b}$ in round $t$ if and only if $\bar{\alpha}$ rejects $\bar{b}'$ in round $t$. From these two facts it follows that the same $M$ and $f$ can be used to satisfy the low relative loss criterion of $\bar{\alpha}_{|N}$ to $\bar{b}$.

□

## A.6 Proof of Theorem 8

**Assumption 1.** *Assume that for each pair of $a_2 \in A_2$ and a best response $a_1^*$ to $a_2$, there is a safe $a_2 \to a_1^*$ bidder who either outbids $\bar{\alpha}$ only finitely many times or whose test set in $\bar{\alpha}$ is weakly uncorrelated with $\bar{\beta}$.*

Theorem 8. *Let $\bar{\alpha}, \bar{\beta}$ be RIAs for the decision problem sequences $\bar{D}^{\alpha}, \bar{D}^{\beta}$, respectively, and no subsequence accounting (i.e., $\mathbb{S}_{\alpha/\beta} = \{\mathbb{N}\}$). If $\alpha^c_t(\mathrm{DP}_t), \beta^c_t(\mathrm{DP}_t)$ converge to choosing with frequency 1, the options corresponding to $a_1 \in A_1, a_2 \in A_2$, then under Assumption 1 for both $\alpha$ relative to $\beta$ and $\beta$ relative to $\alpha$, $(a_1, a_2)$ is a Nash equilibrium of the underlying game.*

Proof. We prove this by contradiction. That is, we assume that $\bar{\alpha}, \bar{\beta}$ converge to choosing some non-NE $(a_1, a_2)$ with frequency 1, and then show a contradiction to the assumption that $\bar{\alpha}, \bar{\beta}$ are RIAs.
WLOG let there be a best response $a_1^* \in A_1$ s.t. $u_1(a_1^*, a_2) > u_1(a_1, a_2)$. Then consider a safe, weakly uncorrelated $a_2 \to a_1^*$ bidder $\bar{b}_i$.
First we show that $\bar{\alpha}$ infinitely often rejects $b_i$. The low absolute loss condition applied to $\bar{\alpha}$ states that

$$\frac{1}{T}\sum_{t=1}^{T} \alpha^e_t(\mathrm{DP}_t) - D_t(\alpha^c_t(\mathrm{DP}_t)) \leq 0 \text{ as } T \to 0. \tag{53}$$

Now by the assumption that $(a_1, a_2)$ is played with limit frequency 1,

$$\frac{1}{T}\sum_{t=1}^{T} D_t(\alpha_t^c(\mathrm{DP}_t)) \to u_1(a_1, a_2) \text{ as } T \to 0. \tag{54}$$

Hence,

$$\frac{1}{T}\sum_{t=1}^{T} \alpha_t^e(\mathrm{DP}_t) \leq u_1(a_1, a_2) \text{ as } T \to 0. \tag{55}$$

It follows in particular that for all $\epsilon > 0$ with positive limit frequency among $t \in \mathbb{N}$, it is

$$\alpha_t^e(\mathrm{DP}_t) < u_1(a_1, a_2) + \epsilon. \tag{56}$$

Because $\beta$ plays $a_2$ with limit frequency 1, $b_i$ (by definition of a safe $a_2 \to a_1^*$ bidder) therefore bids above $u_1(a_1, a_2) + \epsilon$ for some $\epsilon$ with limit frequency 1 and therefore infinitely often outbids $\bar{\alpha}$.
Hence there must be an infinite test set $M$ for $b_i$. As usual, we will assume WLOG (by Lemma 10) that $M$ includes only rounds in which $\bar{b}_i$ submits non-zero bids. Now consider the average relative loss

$$\frac{1}{|M_{\leq T}|} l_T(\bar{\alpha}, \bar{D}^{\alpha}, M, \bar{b}) = \frac{1}{|M_{\leq T}|} \sum_{t \in M_{\leq T}} D_t(b_{i,t}^c(\mathrm{DP}_t)) - \frac{1}{|M_{\leq T}|} \sum_{t \in M_{\leq T}} b_{i,t}^e(\mathrm{DP}_t). \tag{57}$$

By assumption, $\beta$ chooses $a_2$ with limit frequency 1. From this it follows that $\bar{b}_i$ recommends $a_1^*$ with limit frequency 1. By the assumption about weakly uncorrelated testing of $\bar{b}_i$, it also follows that $\beta$ chooses $a_2$ with limit frequency 1 on $M$. From this, it is easy to show that first average converges to $u_1(a_1^*, a_2)$. Since $b_{i,t}^e \leq u_1(a_1^*, a_2) - \epsilon$ for some (constant) $\epsilon > 0$, the second is always less then $u_1(a_1^*, a_2) - \epsilon$. It follows that $\frac{1}{|M_{\leq T}|} l_T(\bar{\alpha}, \bar{D}^{\alpha}, M, \bar{b}) \geq \epsilon$ in the limit and therefore also $l_T(\bar{\alpha}, \bar{D}^{\alpha}, M, \bar{b}) \to +\infty$, violating the low relative loss condition. □

## A.7 Proof of Theorem 9

Theorem 9. *For each game $(A_1, A_2, u_1, u_2)$ and strict Nash equilibrium $(a_1, a_2) \in A_1 \times A_2$ there is a pair of randomizing decision markets $\bar{\alpha}, \bar{\beta}$ that are RIAs with probability 1 relative to any (countable) set of bidders $\mathbb{B}$ (and without subsequence accounting) and that converge to playing $(a_1, a_2)$ with probability 1.*

Proof. We construct the RIAs as follows. Basically we use the same construction as that in Appendix A.2 for the special case of $\mathbb{S} = \{\mathbb{N}\}$. However, we add onto this that in every round with some fixed probability $p \in (0, 1)$, the market chooses the equilibrium action $a_i$ regardless of the highest bidder's recommendation. The estimate in these rounds is nonetheless that of the highest bidder. In these replacement rounds, no bidder is tested and therefore no bidder spends allowance money. The constant $p$ is picked in such a way that it is ensured that the unique best response to this market is always the other player's equilibrium action $a_{-i}$.
We have to show that decision markets constructed in this way are indeed RIAs with probability 1 and that they almost surely converge to playing the given equilibrium $(a_1, a_2)$ with frequency 1.
Low absolute loss: We have to show that per-round absolute loss goes to 0 with probability 1. Let $R$ be the (i.i.d. randomly selected) set of rounds in which $a_i$ is played by "replacement" without any testing. It is

$$\frac{1}{T}\mathcal{L}_T(\bar{\alpha}, \bar{D}, \mathbb{N}) \tag{58}$$

$$= \frac{1}{T}\sum_{T=1}^{T} \alpha_t^e(\mathrm{DP}_t) - D_t(\alpha_t^c(\mathrm{DP}_t)) \tag{59}$$

$$\begin{aligned} &= \frac{1}{T} \sum_{t \in R_{\leq T}} b_{i_t^*,t}^e(\mathrm{DP}_t) - D_t(a_i) \\ &\quad + \frac{1}{T} \sum_{t \in \{1,\ldots,T\} - R} b_{i_t^*,t}^e(\mathrm{DP}_t) - D_t(b_{i_t^*,t}^c(\mathrm{DP}_t)) \end{aligned} \tag{60}$$

$$\begin{aligned} &\leq \frac{1}{T} \sum_{t \in R_{\leq T}} b_{i_t^*,t}^e(\mathrm{DP}_t) - D_t(b_{i_t^*,t}^c(\mathrm{DP}_t) \\ &\quad + \frac{1}{T} \sum_{t \in \{1,\ldots,T\} - R} b_{i_t^*,t}^e(\mathrm{DP}_t) - D_t(b_{i_t^*,t}^c(\mathrm{DP}_t)) \text{ w.p. 1 as } T \to \infty. \end{aligned} \tag{61}$$

The last step is due to the fact that by construction, $a_i$ is always optimal in expectation. Now, the first of the two summands in the last line can be shown to approach 0 by the same argument that we used in the proof of low absolute loss of our RIA algorithm in Appendix A.2: the cumulative loss is bound by allowance distributed (plus the negligible initial wealth) and per-round-allowance goes to 0. But now notice that the second and first sums are the same, except that they are over complementary sets. However, since $R$ is randomly sampled, the terms must approach each other on average, as follows:

$$\begin{aligned}\frac{1}{Tp}\sum_{t\in R_{\leq T}} b^e_{i^*_t,t}(\mathrm{DP}_t) - D_t(b^c_{i^*_t,t}(\mathrm{DP}_t)\\ - \frac{1}{T(1-p)}\sum_{t\in\{1,\dots,T\}-R} b^e_{i^*_t,t}(\mathrm{DP}_t) - D_t(b^c_{i^*_t,t}(\mathrm{DP}_t))\\ \to 0 \text{ w.p. 1 as } T\to\infty.\end{aligned} \tag{62}$$

Hence, because the latter sum is bound by allowance, the former sum is in the limit almost surely bounded by allowance times $1/p$. We conclude that both summands in line 61 approach 0 and therefore that the low absolute loss condition is satisfied.

Low relative loss: The low *relative* loss property can be shown in the same way as in the proof of Theorem 1 in Appendix A.2: whenever a bidder strictly overbids $\alpha$, it must by construction of $\bar{\alpha}$ have insufficient wealth. This in turn implies by how wealth in the construction works that the bidder must have empirically underperformed its estimates.

Convergence to $(a_1, a_2)$: Finally, we need to prove that these RIAs indeed almost surely converge to playing $(a_1, a_2)$ with frequency 1, i.e., that each player plays $a_i$ with frequency 1 rather than just with frequency $p$. This can be shown by essentially the same argument as the proof of Theorem 4 in Appendix A.4. By choice of $p$, recommending $a_i$ guarantees an expected value that is greater than that of any other action. □

# B WHY SUBSEQUENCE ACCOUNTING MAKES A DIFFERENCE

In this section, we give two examples of decision problem sequences $\bar{D}$ in which it seems necessary to let RIAs evaluate bidders on particular subsequences, rather than merely keeping track of their overall performance. Formally speaking, why can we not simply always set $\mathbb{S} = \{\mathbb{N}\}$, at least in the low relative loss criterion, and thereby make our definition much simpler? Of course, there might be solutions other than subsequence accounting for solving these issues, but subsequence accounting is the most elegant solution we have come up with.

## B.1 Subsidizing irrational bidders

For simplicity, we consider the case where $\bar{\alpha}$ is constructed via a decision auction roughly as described in Appendix A.2. Imagine that $\bar{D}$ consists of two types of equally frequent problems, which we call type A and type B problems. Type A can be selected almost arbitrarily. For concreteness, assume that each type A problem is simply {"0.3", "0.1"}, which $\bar{D}$ evaluates in the obvious way. The type $B$ problem in round $t$ consists of a single option $x_t$, whose payoff behavior we will describe in a moment. Let $\bar{b}$ be a particular bidder who bids 0.5 on both type A and type B problems and who recommends selecting "0.1" in type A problems. Let $D_t(x_t)$ be 1 if $\bar{b}$ is the winning bidder in $\bar{\alpha}$ in round $t$ and 0 otherwise.

If $\bar{\alpha}$ does not account separately for subsequences (i.e., if $\mathbb{S} = \{\mathbb{N}\}$ and $b \in \mathbb{B}$), then it is easy to see that $\bar{\alpha}$ will converge to choosing "0.1" and estimating 0.5 on type A problems: overall (i.e., on the entire decision problem sequence), bidder $\bar{b}$ underpromises, thereby amassing wealth. No other bidder can afford to consistently bid as much as $\bar{b}$. This behavior of $\bar{\alpha}$ is not necessarily irrational, even from a myopic perspective. For example, at time $t$ before looking at $\mathrm{DP}_t$, an agent will wish they are the kind that follows $\bar{b}$.

Nonetheless, $\bar{\alpha}$'s poor performance on type A problems is disconcerting. From a pragmatic perspective, it is problematic to have the behavior in type A problems depend on whether our decision problem sequence contains type B problems. If we train an agent with type B decision problems, then this will cause the agent's performance on type A problems to deteriorate. We may then worry that in the real world the agent will face type A problems much more often and that adding type B problems to the training sequence could therefore worsen the agent's real-world performance.

From a more theoretical perspective, it seems inconvenient that type B-like decision problems mess with any guarantees we might hope to obtain on other kinds of decision problems. For example, it seems desirable to have a result like Theorem 3, which (for RIAs with subsequence accounting) guarantees that when one option gives a reward of 0.3 with certainty, at least 0.3 will be obtained. Without subsequence accounting,

any result about behavior on some subsequence of problems would have to predicated on the absence of type B-like problems.

## B.2 Sparse subsequences

We here describe a second, separate problem that is solved by subsequence accounting: poor performance on sufficiently sparse subsequences. Again, we focus somewhat on discussing this problem in the context of the specific algorithm we describe in Appendix A.2.

Remember that our low relative loss criterion for $\mathbb{S} = \{\mathbb{N}\}$ requires that for all bidders $\bar{b}_i$ that outbid $\bar{\alpha}$ infinitely often, the bidder's test set $M_i$ must satisfy

$$\sum_{t \in M_i : t < T} b^e_{i,t}(\mathrm{DP}_t) - D_t(b^c_{i,t}(\mathrm{DP}_t)) \to \infty \tag{63}$$

as $T \to \infty$. Essentially, this means that $\alpha$ must have some tolerance for $\bar{b}_i$'s overpromising. Even if $\bar{b}_i$ overbids somewhat, as long as it remains below the tolerance limit, a RIA has to follow that bidder's recommendation. This tolerance must go to $\infty$. In our decision auction construction of Appendix A.2, tolerance is determined directly by $\bar{b}_i$'s allowance function.

The tolerance can be made to grow arbitrarily slowly (potentially at the cost of convergence rates), as long as each bidder's tolerance goes to $\infty$. Still, for any particular tolerance schedule, the following issue arises. There may be a bidder $\bar{b}_i$ who mostly makes sensible bids, but on a sparse infinite e.d. subsequence overbids and makes poor recommendations. By a "sparse subsequence", we here mean one whose frequency goes to zero faster than the average allowance/tolerance per time step. The overall extent to which the bidder overpromises (the left-hand side of line 63) then grows slower than the tolerance. Hence, the RIA would follow that bidder's recommendations and perform poorly on this subsequence.

As with the sequence described in the previous section, it is not clear whether this is always irrational, especially if we do let the allowance grow very slowly and $\bar{D}$ is somewhat chaotic. After all, rejecting $\bar{b}_i$'s recommendation would be based on a sparse subsequence of the problems so far, and contrary to the perhaps vast number of problems on which the bidder has performed well. On the other hand, there may be cases in which it is intuitive to distrust such a bidder, especially if $\bar{D}$ follows a simple pattern such as provability in Peano arithmetic. For instance, imagine a bidder who at any time $t$ bids using the assumption that any number explicitly represented in $\mathrm{DP}_t$ is not equal to $t^t$. This assumption will almost always be satisfied but lead to obviously suboptimal choices in some rounds. Also, the same theoretical and practical issues as discussed in the previous section apply.

# C MORE ON RANDOMIZATION AND REGRET

In the literature on multi-armed bandit problems, authors usually consider the goal of regret minimization. For any given agent $c$, the Simplified Adversarial Offer $\mathrm{SAO}_c$ of Section 3.2 is a problem on which regret is necessarily high. However, if we assume that the agent at time $t$ can randomize in a way that is independent of $D_t$, it can actually ensured that per-round regret (relative to any particular bidder) goes to 0 (see Section 10).

In this section, we discuss the goal of regret minimization under the assumption that algorithms can randomize independently of $\bar{D}$. The problems discussed in this section all involve references to the agent's choice. In the literature on such Newcomb-like problems (see Section 10), an idea closely related to regret minimization has been discussed under the name ratificationism [see 20, for an introduction and overview].

We consider a version of Newcomb's problem (introduced by [14]; see Section 10 for further discussion and references). In particular, we consider for any chooser $c$ the decision problem $\mathrm{NP}_c = \{a_1, a_2\}$ which is resolved as follows. First, it is

$$D(a_1) = \frac{1}{4} + \frac{1}{2} P(c(\mathrm{NP}_c) = a_1) \tag{64}$$

So the value of $a_1$ is proportional to the probability that $c$ chooses $a_1$. And second, we let $D(a_2) = D(a_1) + P(c(\mathrm{NP}_c) = a_1)/4$.

If we let $p = P(c(\mathrm{NP}_c) = a_1)$, then the expected reward of $c$ in this decision problem is

$$\frac{1}{4} + \frac{1}{2} p + (1 - p)p/4. \tag{65}$$

It is easy to see that this is strictly increasing in $p$ and therefore maximized if $c(\mathrm{NP}_c) = a_1$ deterministically. The regret, on the other hand, of $c$ is $p^2/4$, which is also strictly increasing in $p$ on $[0, 1]$ and therefore

minimized if $c(\text{NP}_c) = a_2$ deterministically. Similarly, the competitive ratio is given by

$$\frac{1/4 + 3p/4}{1/4 + p/2 + (1-p)p/4}, \tag{66}$$

which is also strictly increasing in $p$ on $[0, 1]$ and therefore also minimized if $c(\text{NP}_c) = a_2$ deterministically. Regret and competitive ratio minimization as rationality criteria would therefore require choosing the policy that minimizes the actual reward obtained in this scenario, only to minimize the value of actions not taken. As noted in Section 10, it is a controversial among decision theorists what the rational choice in Newcomb's problem is. However, from the perspective of this paper in this particular version of the problem, it seems undesirable to require reward minimization. Also, it is easy to construct other (perhaps more convincing) cases. For example, if a high reward can be obtained by taking some action with a small probability, then regret minimizers take that action with high probability in a positive-frequency fraction of the rounds. Or consider a version of Newcomb's problem in which $D(a_1)$ is defined as before, but $D(a_2) = D(a_1)$. On such problems, Hannan-consistency is trivially satisfied by any learner, even though taking $a_1$ with probability 1 is clearly optimal.

## D SOME REGRET MINIMIZERS SATISFY A GENERALIZED RIA CRITERION

We here show that some regret minimizers satisfy a slightly generalized version of the RIA criterion. We first have to give a formal definition of regret. Since the literature on adversarial bandit problems with expert advice does not consider experts who submit estimates in the way that our bidders do, we cannot use an existing definition and will instead make up our own. For simplicity, we will only consider the case $\mathbb{S} = \mathbb{N}$. Let $\bar{D}$ be a decision process, $\bar{\alpha}$ be a decision market and $\mathbb{B} = \{b_1, b_2, ...\}$ be a set of bidders. For simplicity, let $\mathbb{B}$ be finite. For each $b_i \in \mathbb{B}$, let $B_i := \{t \in \mathbb{N} \mid b^e_{i,t}(\text{DP}_t) > \alpha^e_t(\text{DP}_t)\}$ be the set of rounds in which $b_i$ outbids $\alpha$. We define the average per-round regret of the learner to bidder $b_i$ up to time $T$ as

$$\text{REGRET}_{m,T} = \mathbb{E}\left[\frac{1}{|B_{m,\leq T}|}\sum_{t \in B_{m,\leq T}} D_t(b^c_{i,t}(\text{DP}_t)) - \alpha^e_t(\text{DP}_t)\right]. \tag{67}$$

As before, the bidding mechanisms means that bidders can specialize on specific types of decisions.[3] As is common in the adversarial bandit problem literature, we will be interested in learning algorithms that guarantee average regret to go to zero as $|B_{m,\leq T}| \to \infty$.

Regret is somewhat analogous to relative loss. As with relative loss, low regret can be achieved trivially by setting $\alpha^e = 1$. Thus, if we replace the low relative loss requirement with a sublinear regret requirement, we have to keep the low *absolute* loss assumption.

**Conjecture 12.** *Let $\bar{D}$ be a decision process where $|\text{DP}_t|$ is bounded for all $t \in \text{N}$. With access to an independent source of randomization, and given access to the outputs of all bidders in $\mathbb{B}$, we can compute $\bar{\alpha}$ with low absolute loss on $\mathbb{N}$ s.t. for all bidders $b_i$, $\text{REGRET}_{i,T} \to 0$ with probability 1 if $|B_{i,\leq T}| \to \infty$.*

As noted elsewhere, without independent randomization it is clear that such an $\bar{\alpha}$ cannot be designed. Even with independent randomization, it is not obvious whether the conjecture holds. However, similar results in the literature on adversarial bandit problems with expert advice lead us to believe that it does. That said, we have not been able to prove the conjecture by using simply the results from that literature.

Theorem 13. *Let $\bar{\alpha}$ be an independently randomized decision market that has low absolute loss on $\mathbb{N}$ and ensures sublinear regret with probability 1 relative to all bidders in some finite set $\mathbb{B} = \{b_i\}_i$. Further assume that for all bidders $b_i$, $P(\alpha^c_t(\text{DP}_t){=}b^c_{i,t}(\text{DP}_t)) \in \omega(1/t)$ among $t \in B_i$. Then we can compute based on $\alpha$ a new decision market $\tilde{\alpha}$ that has low absolute loss and that satisfies for each bidder $b_i$ that is infinitely often rejected,*

$$\sum_{t \in B_{i,\leq T}} \frac{\mathbb{1}[\alpha^c_t(\text{DP}_t){=}b^c_{i,t}(\text{DP}_t)]}{P(\alpha^c_t(\text{DP}_t){=}b^c_{i,t}(\text{DP}_t))}(D_t(b^c_{i,t}(\text{DP}_t)) - b^e_{i,t}(\text{DP}_t)) \to -\infty \tag{68}$$

*among $T$ at which $\tilde{\alpha}$ rejects $b_i$.*

Notice that the left-hand side of line 68 is a weighted version of the relative loss on the set $\{t \in B_{i,\leq T} \mid \alpha^c_t(\text{DP}_t){=}b^c_{i,t}(\text{DP}_t)\}$.

[3] Note that we subtract the decision market's *estimates*, not the utility that $\bar{\alpha}$ in fact achieves. This is important. Otherwise, the learner can set $\alpha^e = 0$ even in rounds in which $D_t(\alpha^c_t(\text{DP}_t))$ is (expected to be) high, thus circumventing the expert's bidding mechanism.

Still, there are alternative definitions that also work. For example, one might count regret only in rounds in which $\alpha$ and $b_i$ differ in their recommendations.

The proof combines one key idea from the literature on adversarial multi-armed bandits – importance-weighted estimation – and one from this paper – the decision auction construction (Appendix A.2).

Proof. For $t \in B_i$, define

$$\hat{\mathrm{R}}_{i,t} = \frac{\mathbb{1}[\alpha_t^c(\mathrm{DP}_t)=b_{i,t}^c(\mathrm{DP}_t)]}{P(\alpha_t^c(\mathrm{DP}_t)=b_{i,t}^c(\mathrm{DP}_t))}(D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t)),$$

where we assume $P(\alpha_t^c(\mathrm{DP}_t)=b_{i,t}^c(\mathrm{DP}_t)) > 0$. As usual it is then,

$$\mathbb{E}\left[\hat{\mathrm{R}}_{i,t}\right] = D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t).$$

For $t \notin B_i$, define $\hat{\mathrm{R}}_{i,t} = 0$. Hence, $\hat{\mathrm{R}}_{i,t}$ can be used as an unbiased estimator of the regret in a single round. Further, $\mathrm{Var}(\hat{\mathrm{R}}_{m,t}) \in o(t)$, and thus $\sum_{t=1}^{T} \mathrm{Var}(\hat{\mathrm{R}}_{m,t}) \in o(T^2)$. By Kolmogorov's strong law of large numbers,

$$\frac{1}{T}\sum_{t\in B_{m,\leq T}} \hat{\mathrm{R}}_{m,t} - \frac{1}{T}\sum_{t\in B_{m,\leq T}} D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t) \tag{69}$$

$$= \quad \frac{1}{T}\sum_{t=1}^{T} \hat{\mathrm{R}}_{m,t} - \frac{1}{T}\sum_{t=1}^{T} D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t) \tag{70}$$

$$\to \quad 0 \text{ as } T \to \infty \tag{71}$$

In other terms,

$$\sum_{t\in B_{m,\leq T}} \hat{\mathrm{R}}_{m,t} - \sum_{t\in B_{m,\leq T}} D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t)$$

is sublinear.

We now construct new estimates. Fix a non-decreasing, sublinear function $\mathrm{CA}\colon \mathbb{N} \to \mathbb{R}$ with $\mathrm{CA}(T) \to \infty$. (These are cumulative versions of the allowance functions from the construction in Appendix A.2.) Next, we define

$$\begin{aligned}\mathcal{L}_{i,T} \coloneqq & \sum_{t\in M_{i,\leq T}} \frac{\mathbb{1}[\alpha_t^c(\mathrm{DP}_t)=b_{i,t}^c(\mathrm{DP}_t)]}{P(\alpha_t^c(\mathrm{DP}_t)=b_{i,t}^c(\mathrm{DP}_t))}(D_t(b_{i,t}^c(\mathrm{DP}_t)) - b_{i,t}^e(\mathrm{DP}_t)) \\ & + \sum_{t\in B_{i,\leq T}-M_i} \frac{\mathbb{1}[\alpha_t^c(\mathrm{DP}_t)=b_{i,t}^c(\mathrm{DP}_t)]}{P(\alpha_t^c(\mathrm{DP}_t)=b_{i,t}^c(\mathrm{DP}_t))}(D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t)),\end{aligned}$$

where $M_i \subseteq B_i$ will be defined in a second. Define $w_T(i) = \mathrm{CA}(T) + \mathcal{L}_{i,T}$. Now at each time $t$, we define our new estimate as

$$\tilde{\alpha}_t^e(\mathrm{DP}_t) = \max(\alpha_t^e(\mathrm{DP}_t), \max_{i:w_{t-1}(i)\geq 0} b_{i,t}^e(\mathrm{DP}_t)). \tag{72}$$

Finally, let $M_i$ be the set of rounds in which $i$ is the maximizer in Eq. 72 through the outer max.

We now need to show two things: That absolute loss is still sublinear even for the new increased $\tilde{\alpha}_t^e(\mathrm{DP}_t)$ and that the claimed low relative loss condition is satisfied.

We start with the low relative loss part. First notice that because $M_i \subseteq B_i$ and for $t \in B_i$, $b_{i,t}^e(\mathrm{DP}_t) > \alpha_t^e(\mathrm{DP}_t)$, we get that

$$w_T(i) \geq \mathrm{CA}(T) + \sum_{t\in B_{m,\leq T}} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{P_t(\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t))}(D_t(b_{i,t}^c(\mathrm{DP}_t)) - b_{i,t}^e(\mathrm{DP}_t)).$$

Thus, whenever $b_{i,T}^e(\mathrm{DP}_t) > \tilde{\alpha}_T^e(\mathrm{DP}_t)$, then by construction $w_t(i) < 0$, and therefore

$$\sum_{t\in B_{i,\leq T}} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{P(\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t))}(D_t(b_{i,t}^c(\mathrm{DP}_t)) - b_{i,t}^e(\mathrm{DP}_t)) \leq -\mathrm{CA}(T).$$

Thus, we get that among $T \in \tilde{B}_i$ (the times where $t$ strictly outbids the new estimates), the empirical record on the test set goes to $-\infty$.

It is left to show that absolute loss remains low if we increase the estimates from $\alpha^e$ to $\tilde{\alpha}^e$. We have

$$\sum_{t=1}^{T} \tilde{\alpha}_t^e(\mathrm{DP}_t) - D_t(b_{i,t}^c(\mathrm{DP}_t)) = \sum_{t=1}^{T} \alpha_t^e(\mathrm{DP}_t) - D_t(b_{i,t}^c(\mathrm{DP}_t)) + \sum_{t=1}^{T} \tilde{\alpha}_t^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t).$$

The first sum is sublinear by assumption. So we only have to show that $\sum_{t=1}^{T} \tilde{\alpha}_t^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t)$ is sublinear in $T$. We have

$$\sum_{t=1}^{T} \tilde{\alpha}_t^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t) = \sum_i \sum_{t \in M_{i,\leq T}} b_{i,t}^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t). \tag{73}$$

So, it is left to show that the increase on behalf of each expert $i$ is sublinear.
Now, we use IWE again. That is, we consider

$$\sum_{t \in M_{i,\leq T}} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{P(\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t))} (b_{i,t}^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t)).$$

By the same argument as above, we can show that the difference between this term and $\sum_{t \in M_{i,\leq T}} b_{i,t}^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t)$ is sublinear. So it is enough to show that this term is sublinear.
Now notice that

$$\begin{aligned}
w_T(i) &= \mathrm{CA}(T) + \sum_{t \in M_{i,\leq T}} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{P(\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t))} \underbrace{(D_t(b_{i,t}^c(\mathrm{DP}_t)) - b_{i,t}^e(\mathrm{DP}_t))}_{=(D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t)) - (b_{i,t}^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t))} \\
&\quad + \sum_{t \in B_{i,\leq T} - M_i} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)} \left(D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t)\right) \\
&= \mathrm{CA}(T) - \sum_{t \in M_{i,\leq T}} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{P_t(\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t))} (b_{i,t}^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t)) \\
&\quad + \sum_{t \in B_{i,\leq T}} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{P_t(\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t))} (D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t)).
\end{aligned}$$

Now, for $T \in M_i$, it must be $w_T(i) \geq 0$. Still, $w_T(i)$ can fall under 0, but only by $\hat{R}_t^m$ for some $t \in \{1, ..., T\}$, which is in $o(T)$. Thus,

$$\begin{aligned}
&\sum_{t \in M_{i,\leq T}} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{P_t(\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t))} (b_{i,t}^e(\mathrm{DP}_t) - \alpha_t^e(\mathrm{DP}_t)) \\
\leq\ &\mathrm{CA}(T) + \sum_{t \in B_{i,\leq T}} \frac{\mathbb{1}\left[\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t)\right]}{P_t(\alpha_t^c(\mathrm{DP}_t) = b_{i,t}^c(\mathrm{DP}_t))} (D_t(b_{i,t}^c(\mathrm{DP}_t)) - \alpha_t^e(\mathrm{DP}_t)) + o(T)
\end{aligned}$$

CA is sublinear by construction and the second summand has been shown to be sublinear above. □