Relevant deviation rules in rationalizing dynamic choices*

Henrique de Oliveira Rohit Lamba

April 2025

Preliminary

Abstract

An agent takes a sequence of actions. The analyst does not have access to the agent's information but knowns the utility function. Assuming the agent is Bayesian and an expected utility maximizer, de Oliveira and Lamba [2025] established that the observed actions cannot be justified if and only if there is a single deviation argument that leaves the agent better off, regardless of the information. This paper develops an efficient method to characterize the set of such deviation arguments. Specifically, the approach reduces the number of potentially binding deviation rules that need to be considered, thereby enhancing computational tractability.

1 Introduction

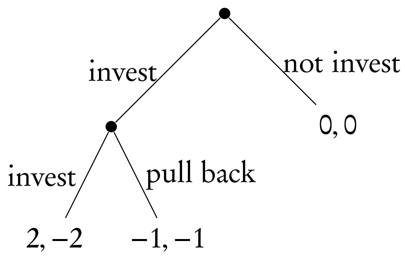
Suppose an agent takes a sequence of actions and observes a potentially new piece of information about a payoff relevant state before taking each action. A natural question to ask is: What action sequences can an outside analyst who understands the agent's preferences and observes these actions but not the information rationalize? A few examples make the question concrete.[1]

Example 1. A CEO faces an opportunity to invest in a project with uncertain payoffs: there is a return of 4 if the project meets favorable conditions in the future (good state) and 0 if not (bad state). The project bears fruits on two rounds of investment, and each round of investment costs 1 unit. The CEO has three options: not invest, invest in the first round and pull back in the second, or investment in both periods. The CEO's payoff matrix and decision tree can be summarized as follows:

*de Oliveira: Fundação Getúlio Vargas-EESP, henrique.deoliveira@fgv.br; Lamba: Cornell University, rohitlamba@cornell.edu We are grateful to Gabriel Carroll and Ashwin Kambhampati for helpful comments.

[1] We invoked these examples first in our work de Oliveira and Lamba [2025].

	not invest	invest & pull back	invest & invest
good	0	-1	2
bad	0	-1	-2

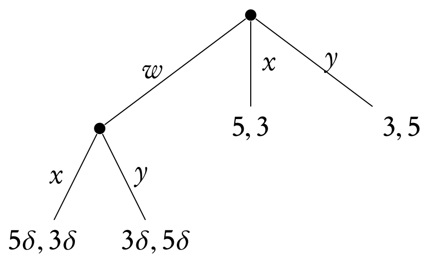


Suppose that we learn that the CEO invested in the first round, incurring the initial cost, but then pulled back. Some might interpret that as evidence of incompetence, saying that in no state can this sequence of actions be justified. They might say that even if the CEO was not sure about the state of the world, not investing would surely have been a better choice. These critics would be ignoring a simple explanation: it might be that the CEO initially received good news about the investment, but after the first round of investment learnt that the project was likely to fail.

In Example 1, the action sequence (*invest*, *pull back*) is what we will call *apparently dominated*—there exists another sequence of actions, (*not invest*, $\emptyset$), under which the agent does strictly better in every state of the world.[2] It will be easy to show that any action sequence that cannot be rationalized is apparently dominated. However, as Example 1 shows, the converse is not true. In fact, in Example 1, all three possible sequences of choices can be rationalized, which illustrates how permissive this first criterion is. But it is not vacuous and can exclude some dynamic choices. For instance, consider the following example:

Example 2. A firm can bet on one of two technologies, X or Y. The firm can also postpone the decision, but by doing so its payoff is discounted by a factor δ, where $0 < \delta < 1$. The payoff matrix and decision tree are as follows:

	x	y	wx	wy
X	5	3	5δ	3δ
Y	3	5	3δ	5δ



Note that both wx and wy are apparently dominated, which does not necessarily rule them out.[3] We learn that the firm has decided to wait instead of making an immediate bet. Under what values of δ can this choice be rationalized? By waiting, the firm can get at most 5δ. By

[2] Generally, an action sequence is apparently dominated if there exists another action sequence (or a lottery over action sequences) that does strictly better in every state of the world.

[3] We are using the shorthand wx for (w, x) and wy for (w, y).

making an immediate decision, the firm is guaranteed to get at least 3. Hence, if $\delta < 3/5$, waiting cannot be rationalized.

But this is not the full story. If the firm makes an immediate decision to randomize equally between x and y, it is guaranteed an expected payoff of 4, no matter the state. Therefore, waiting cannot be rationalized when $\delta < 4/5$. On the other hand, if $\delta \geqslant 4/5$, waiting can be explained by the following information: it could be that the firm starts with an even prior and then fully learns the state of the world in the second period. Thus, waiting can be rationalized precisely when $\delta \geqslant 4/5$.

More generally, for an arbitrary T-period decision problem, an action sequence can be rationalized when there exists a prior and a sequential information structure for which an optimizing agent could end up choosing that action sequence with positive probability. Thus, to argue that an action sequence can be rationalized, it is enough to provide a single information structure and prior that prove it to be so; to argue that an action sequence cannot be rationalized, it must be shown that every information structure and prior would fail to rationalize it.

In de Oliveira and Lamba [2025], we characterize the empirical content of this model– more specifically, we present a dominance argument that provides a dual characterization of *when an action sequence cannot be rationalized by some sequential information structure*. Generally speaking, the argument is based on the idea of a deviation rule, which prescribes the agent to deviate in specific ways while respecting the structure of the decision tree. If such a deviation leads to a strict improvement in payoff with respect to the action sequence under consideration and does not worsen payoffs elsewhere on the tree, we have dominance that ensures the said action rule cannot be justified. For example, in Example 2, we found a single deviation that simultaneously showed that every information structure would fail to rationalize waiting, thus avoiding direct consideration of the set of all information structures.

Formally, a deviation rule is an *adapted* mapping from action sequences to lotteries over action sequences, $D : A \to \Delta(A)$. Adaptedness simply requires that deviations today can only be a function of past actions and past deviations, and not of future actions or deviations. In Example 1, if we map (*invest*, *pull back*) to (*not invest*, $\emptyset$), then adaptedness demands that we have to map (*invest*, *invest*) also to (*not invest*, $\emptyset$).

Now, in full generality, as the size of the decision problem increases, so will the set of all possible deviation rules. Although in many problems of economic interest, the appropriate deviation rule that can potentially establish true dominance may be intuitive, but from the perspective of computational complexity, it is useful to have a way of dealing with the constraint set imposed by understanding rationalization through deviation rules. In fact, in de Oliveira and Lamba [2025], we show how the search for binding deviation rules can be reduced to an intuitive linear proogram. Is there a formal way to reduce the size of the constraint set in this program?

In this paper, we devise such a way to restrict the class of potentially binding deviation rules. The analysis is based on two simple ideas: If an action sequence is indeed dominated in the sense described above, it seems wasteful to recommend deviating from the tree that contains that action sequence. In addition, it is also unnecessary to form chains of deviations: if it makes sense to deviate towards some action sequence, say **b**, then perhaps it also makes sense not to deviate away from it.

The main result in the paper formalizes how the incorporation of these two observations significantly reduces the set of potentially binding pure deviation rules, thereby reducing the complexity of the underlying linear program. For example, in the context of Example 1, the number of *relevant* pure deviation rules goes from a total of twenty seven to merely two.

Now, deviation rules are "mixed" strategies, and the randomization in an important part of why they deliver a duality result for rationalizing dynamic choices through true dominance. However, our result delivers a shrinking of the set of "pure" deviation rules. How should we think about the gap?

One way we think about this problem is that the set of deviation rules is a polytope in $\mathbb{R}^{A\times A}$, and the problem of finding a dominating deviation rule reduces to a linear program involving this polytope.[4] Thus, it is useful to have a good description of this polytope.

A natural way to describe a polytope is via its extreme points. The extreme points of the polytope of deviation rules are always pure deviation rules. This is why pure deviation rules are important. When we say that "only two of them are relevant", we mean that every relevant mixed deviation rule can be written as a mixture of those two pure deviation rules. This description is usually very helpful for finding the mixed deviation rules.

2 Model and definitions

2.1 Notation

A *stochastic map* from X to a finite set Y is a function $\alpha : X \to \Delta(Y)$, where $\Delta(Y)$ is the set of probability distributions over Y. We represent the probability assigned to y at the point x by

[4] This is established in Proposition 4 in de Oliveira and Lamba [2025]

$\alpha(y|x)$. The composition of two stochastic maps $\alpha : X \to \Delta(Y)$ and $\beta : Y \to \Delta(Z)$ is given by

$$\beta \circ \alpha(z|x) = \sum_{y \in Y} \beta(z|y)\alpha(y|x).$$

We can think of a lottery as a stochastic mapping whose domain is a singleton. Therefore, given $\alpha \in \Delta(Y)$ and $\beta : Y \to \Delta(Z)$, we write

$$\beta \circ \alpha(z) = \sum_{y \in Y} \beta(z|y)\alpha(y)$$

to be the probability with which z is chosen by $\beta \circ \alpha$.

For a real-valued function $u : Y \to \mathbb{R}$ and for a lottery $\alpha \in \Delta(Y)$, we denote by $u(\alpha) = \sum_{y \in Y} \alpha(y)u(y)$ the expected value of $u(\cdot)$ under the distribution α.

Throughout the text, we consider a finite number of time periods $t = 1, \ldots, T$. For a collection of sets $(X_t)_{t=1}^T$, we will use the following notation

$$X^t = \prod_{\tau=1}^{t} X_\tau \qquad X = \prod_{\tau=1}^{T} X_\tau$$

with elements $\mathbf{x}^t \in X^t$ and $\mathbf{x} \in X$. Finally, a stochastic map $\alpha : X \to \Delta(Y)$ is said to be *adapted* if the marginal probability of the first t terms of $\mathbf{y}$ depends only on the first t terms of $\mathbf{x}$; formally, it is adapted if the function

$$\sum_{y_{t+1}, \ldots, y_T} \alpha(y_1, \ldots, y_t, y_{t+1}, \ldots y_T | x_1, \ldots, x_t, x_{t+1}, \ldots, x_T)$$

is constant in $x_{t+1}, \ldots, x_T$.

2.2 The model

The formal objective is to characterize the empirical content of the joint hypothesis of (Bayesian) rationality and the specific payoff function, but without any hypothesis on the information structure. With that objective in mind, we introduce the formal model. The decision problem is shown in Figure 1.

In each time period t, the agent chooses an action a_t from a finite set A_t. Payoffs are determined after period T by a utility function $u(\mathbf{a}, \omega)$, which depends on the entire action sequence $\mathbf{a} = (a_1, \ldots, a_T) \in A$ and a potentially unknown state of the world ω drawn from a finite set Ω.

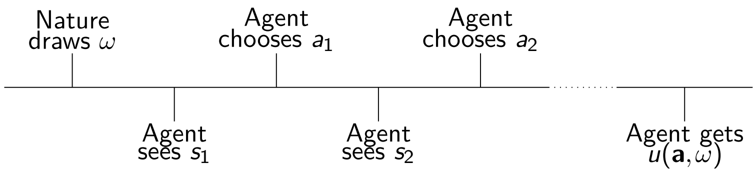


Figure 1: The timeline of signals and actions

There are no other restrictions on the utility function.

The agent is informed about the underlying state of the world over time through a sequence of signals. The timeline of the dynamic decision problem is expressed in Figure 1. Every period, before taking an action, the agent observes a signal that is (potentially) correlated with the state of the world and with the signals she has observed in the past. Formally, the sequence of signals is generated by a *sequential information structure*:

Definition 1. *A* **sequential information structure** *is a sequence of finite sets of signals* $(S_t)_{t=1}^T$ *and a stochastic mapping* $\pi : \Omega \to \Delta(S)$.[5]

We will often refer to the sequential information structure simply as π; the set of signals shall be implicit. The agent's strategy maps each sequence of signals into a lottery over actions every period, with the restriction that the agent cannot base the choice of an action on signals that have not yet been revealed, which we call adaptedness.

Definition 2. *A* **strategy** *for the agent is an adapted stochastic mapping* $\sigma : S \to \Delta(A)$.[6]

Given the sequential information structure π and agent's strategy σ, the probability that the agent takes a given sequence of actions in each state of the world ω is given by $\sigma \circ \pi(\mathbf{a}|\omega)$. Finally, given a prior $p \in \Delta(\Omega)$, she can evaluate her expected payoff:

$$U(\sigma, \pi, p) = \sum_{\omega \in \Omega} p(\omega) \sum_{\mathbf{a} \in A} \sigma \circ \pi(\mathbf{a}|\omega)\, u(\mathbf{a}, \omega).$$

The agent's problem then is to choose an optimal σ given π and p. Throughout the paper, we refer to this model of decision making as the *Bayesian model*.

Our goal is to characterize the empirical content of this model. To that end, we say that an

[5]We can equivalently define the sequential information structure period-by-period as follows. Let $\pi = (\pi_t)_{t=1}^T$ be a family of stochastic mappings where $\pi_1 : \Omega \to \Delta(S_1)$, and $\pi_t : \Omega \times S^{t-1} \to \Delta(S_t)\ \forall\, 2 \le t \le T$. Except for zero probability events, we can deduce that the two definitions are equivalent. The minor distinction does not affect the agent's utility and is therefore irrelevant for our results. For a proof, see Lemma 3 in de Oliveira [2018].

[6]As with information structures, an equivalent way to think of the agent's strategy is a family of stochastic mappings $\sigma = (\sigma_t)_{t=1}^T$, where $\sigma_1 : S_1 \to \Delta(A_1)$, and $\sigma_t : S^t \times A^{t-1} \to \Delta(A_t)\ \forall\, 2 \le t \le T$. It is possible to deduce one formulation from the other.

action sequence can be *rationalized* if it can be chosen with positive probability by an optimizing agent with some information structure and some prior.

Definition 3. *An action sequence* $\mathbf{a} \in A$ *can be* **rationalized** *if there exists a triplet* (σ, π, p) *such that:*

1. $\sigma \in arg \max_{\hat{\sigma}} U(\hat{\sigma}, \pi, p)$ *and*

2. $\sigma \circ \pi \circ p(\mathbf{a}) > 0$.[7]

This definition is permissive in the sense that an action sequence is considered rationalized even if its probability is very small, as long as it is positive. To deduce that an action sequence cannot be rationalized, the analyst needs to work through all possible pairs (π, p), and show that the corresponding optimal strategy σ will not pick that action sequence with positive probability. Since the set of all sequential information structures is quite large, this poses a challenge. The main goal of de Oliveira and Lamba [2025] was to find an alternative way to characterize the set of action sequences that cannot be rationalized, which we now briefly summarize.

3 Rationalizing dynamic choices: a tight characterization

3.1 Deviation rules and true dominance

A *deviation rule* is an adapted mapping $D : A \rightarrow \Delta(A)$, where recollect that being adapted means that the marginal distribution on A^t, the (potentially random) deviation strategy for the first t periods, depends only on A^t, the first t elements of the original strategy from which the agent is deviating. We can think of the deviation rule as a list of alternative actions the agent would take as a function of the actions she originally intended to take. Importantly, a deviation rule is a fully prescribed plan, so if σ is the original strategy, then $D \circ \sigma(\mathbf{a}|\mathbf{s})$ is also a well-defined strategy.

Now, we are in a position to define the appropriate notion of dominance for our model.

Definition 4. *A deviation rule* $D : A \rightarrow \Delta(A)$ **dominates** *an action sequence* $\mathbf{a}$ *if*

1. $u(D(\mathbf{a}), \omega) > u(\mathbf{a}, \omega)$ *for all* $\omega \in \Omega$.

2. $u(D(\mathbf{b}), \omega) \geqslant u(\mathbf{b}, \omega)$ *for all* $\mathbf{b} \in A$ *and* $\omega \in \Omega$.

We say that $\mathbf{a}$ *is* **truly dominated** *if there exists a deviation rule that dominates it.*

[7] Here $\sigma \circ \pi \circ p(\mathbf{a}) = \sum_{\omega} \sigma \circ \pi(\mathbf{a}|\omega)\, \mathbf{p}(\omega)$ (see Section 2.1).

The first part of the definition requires that the action sequence that is to be dominated be strictly improved upon. The second part requires that the payoff induced by the deviation rule do not become worse for any other action sequence in any state. Moreover, there's no visible time dimension in the definition above; time is implicit in the condition that D must be adapted. For $T = 1$, the same definition applies, but the condition that D is adapted becomes vacuous, and so does the second part of the definition. In that case, if a is strictly dominated by α, we can define a deviation rule D_α that takes a to α and does not change any other actions. D_α then dominates a according to the definition above.

When $T > 1$, the adaptedness restriction prevents the construction of such a simple deviation rule—if D specifies a change for the first action in the sequence **a**, then it must specify the same change for all sequences **b** that share the same first action, and so on. The second condition and the embedded notion of adaptedness in the definition impose meaningful restrictions when $T > 1$, encapsulating the distinction between true and apparent dominance.

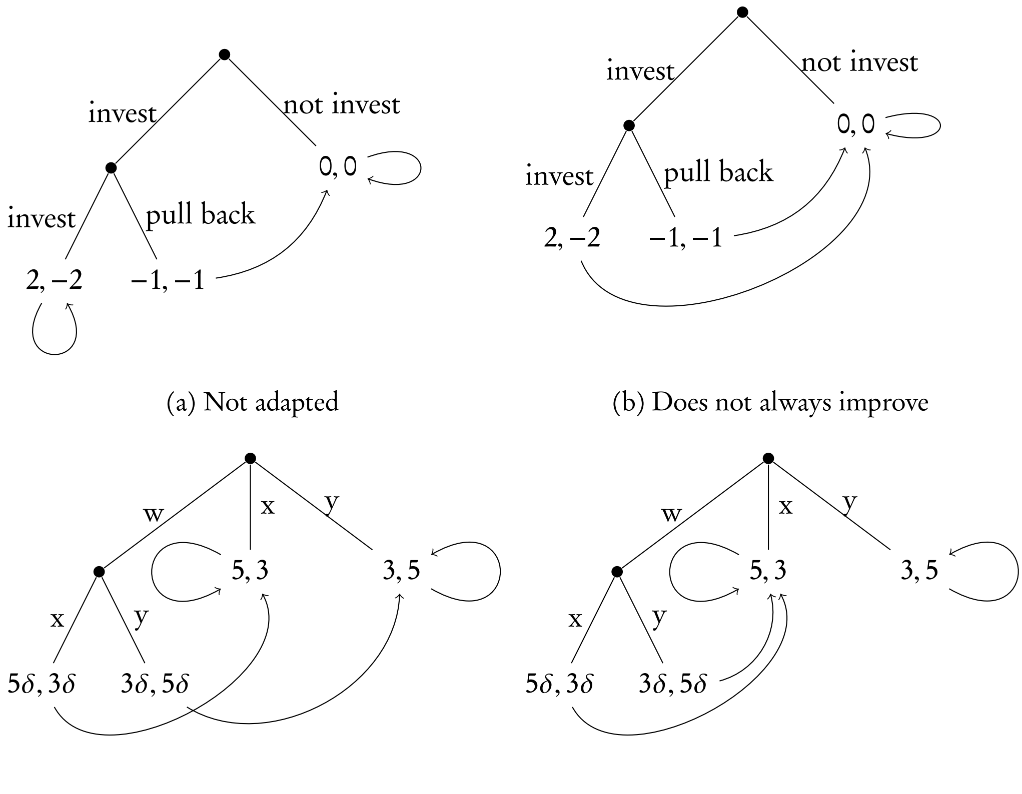


Figure 2: Deviation rules for Examples 1 and 2

To better grasp the definitions of deviation rule and true dominance, Figure 2 illustrates the concepts in the context of our examples. Each complete sequence of actions corresponds to a terminal node. Thus, any mapping from sequences of actions into sequences of actions is depicted as arrows between terminal nodes. It is instructive to note which deviation rules are and which are not adapted, and when the simple adapted ones do not improve upon the intended action sequences.

For instance, in the waiting example, the "deviation rule" depicted in Figure 2c is not adapted, since it represents the infeasible advice "whatever you would choose in the second period, choose the same in the first period". The deviation rule in Figure 2d represents the advice "if you were thinking about waiting, choose x instead", which is adapted. When $\delta < \frac{3}{5}$, it dominates wx and wy, but when $\delta > \frac{3}{5}$ it does not dominate wx nor wy, because x may give a strictly lower payoff than wy. For the tightest possible statement, we therefore constructed the deviation rule $wx \mapsto \frac{1}{2}x + \frac{1}{2}y$, $wy \mapsto \frac{1}{2}x + \frac{1}{2}y$, $x \mapsto y$ and $y \mapsto y$ which (simultaneously) truly dominates wx and wy if and only if $\delta < \frac{4}{5}$.

3.2 Reviewing the characterization result

The main result from de Oliveira and Lamba [2025] is as follows.

Theorem 1. *A sequence of actions cannot be rationalized if and only if it is truly dominated.*

The theorem provides a tight characterization of the set of action sequences that cannot be rationalized. Through its duality formulation, it simplifies their identification by requiring the analyst to construct *one* deviation rule as opposed to treading through the family of *all* sequential information structures.

Now, the set of mappings from all possible action sequences to all possible action sequences can be quite large, even when restricted by adaptedness. Often, the "salient" or "binding" deviation rule is intuitive. But it is also useful to have a systematic way to reduce the set of deviation rules that the analyst may need to look at. In the following main result of this paper, we provide one such method of reducing the set of deviation rules by removing some clearly bad candidates.

4 The main result: Simple deviations

Is there a systematic way to rule out poor candidate deviation rules? Intuitively speaking, if **a** is truly dominated, it seems futile to recommend deviating to the tree which contains **a**. Moreover, it also seems unnecessary to form chains of deviations: if it makes sense to deviate towards some action sequence **b**, then perhaps it also makes sense not to deviate away from it. Here we formalize these intuitions by thinking of a deviation rule as a Markov chain. Recollect that for

$D(\mathbf{b}) \in \Delta(A)$, we refer to $D(\mathbf{a}|\mathbf{b}) \in [0, 1]$ as the weight put on $\mathbf{a}$ by the probability distribution $D(\mathbf{b})$.

Definition 5. *Given a deviation rule D, we say that*

1. $\mathbf{a}$ is *repulsive* if $D(\mathbf{a}|\mathbf{b}) = 0$ for all $\mathbf{b} \in A$;

2. $\mathbf{a}$ is *absorbing* if $D(\mathbf{a}|\mathbf{a}) = 1$.

The term *absorbing* is directly borrowed from the Markov chains taxonomy, and the notion of *repulsiveness* is closely related to the idea of accessibility. In Markov chains, we say a "state" $\mathbf{a}$ is *accessible* from state $\mathbf{b}$ in n steps if $D^n(\mathbf{a}|\mathbf{b}) > 0$, where D^n represents the application of the deviation rule n times. The "state" $\mathbf{a}$ is then said to be *inaccessible* from $\mathbf{b}$ if $D^n(\mathbf{a}|\mathbf{b}) = 0$ for all n. It is easy to see that $\mathbf{a}$ is repulsive if and only if it is inaccessible from all other "states" $\mathbf{b}$. We can now operationalize this terminology to prune the decision tree of truly dominated action sequences.

Definition 6. *A deviation rule D* removes $(a_1, \ldots, a_t)$ *if*

1. $\mathbf{b}$ *is repulsive whenever* $(b_1, \ldots, b_t) = (a_1, \ldots, a_t)$*, and*

2. $\mathbf{b}$ *is absorbing whenever* $(b_1, \ldots, b_t) \neq (a_1, \ldots, a_t)$.

Theorem 2. *If* $\mathbf{a} = (a_1, \ldots, a_T)$ *is a truly dominated action sequence, then there exists a deviation rule D that dominates* $\mathbf{a}$ *and removes* $(a_1, \ldots, a_t)$ *for some* $t \geqslant 1$.

As a thought experiment, let us apply this result to Example 1 in the introduction. Suppose we are trying to show that (*invest*, *pull back*) is truly dominated. There are a total of 15 pure decision rules to consider and, in principle, we would need to look for a dominating deviation rule among all their mixtures.[8] Using the theorem above, we can simplify this search dramatically. First, suppose that D removes $a_1 =$ *invest.* There is a single deviation rule that does that, namely the one that always deviates to *not invest* (see Figure 3a). Now, suppose that D removes $(a_1, a_2) =$ (*invest*, *pull back*). Then (*invest*, *invest*) will be absorbing, and by adaptedness it is the only candidate for a deviation from (*invest*, *pull back*). Therefore, we again have a single candidate for D, namely the one that recommends deviating from (*invest*, *pull back*) to (*invest*, *invest*) and does not recommend any other deviation (see Figure 3b). Thus, the theorem tells us that we only have to consider two deviations.

[8] There are 3 action sequences so a total of 27 possible (pure) deviation mappings, but not all of them are adapted. It can be checked that exactly 15 combinations are possible for mappings that are adapted.

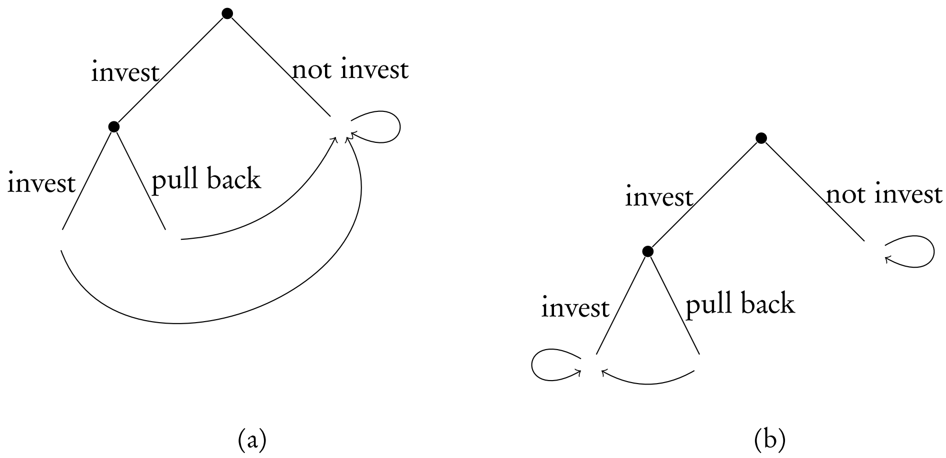


Figure 3: Relevant deviations for the investment example

5 Relevant deviation rules in simple stopping problems

We now apply our result to a class of *simple stopping problems*: In each period, the agent has two possible actions: "stop" and "continue". If the agent stops, the decision tree ends, and the payoff is realized. The decision problem has T periods, so in the final period the tree ends either way. As before, there is an underlying state of the world that can take on at least two values.

The following proposition shows the total number of deviation rules in a stopping problem.

Proposition 1. *The number of pure deviation rules in a T-period stopping problem is*

$$F(T) = \sum_{k=0}^{T} \frac{(T+1)!}{k!}.$$

Proof. We prove this by induction. For $T = 1$, it is easy to see that there are 4 deviation rules, and we have

$$F(1) = \frac{2!}{0!} + \frac{2!}{1!} = 4.$$

Now suppose that we have proven the formula for $T-1$. First, we consider all the deviation rules that deviate from "continue" to "stop" in the first period. The only question then is where we deviate from "stop" in the first period to. There are $T+1$ possible deviations (stopping in each of the T periods or always continuing).

Now we consider deviation rules that do not deviate from "continue" in the first period. There are, as before, $T+1$ ways of deviating from "stop" in the first period, and for each of those, we have $F(T-1)$ ways of continuing deviating for the other sequences. This proves the formula

$$F(T) = T + 1 + (T+1)F(T-1).$$

Substituting our formula for $T-1$, we get

$$\begin{aligned} F(T) &= (T+1) + (T+1)\sum_{k=0}^{T-1} \frac{T!}{k!} \\ &= \frac{(T+1)!}{T!} + \sum_{k=0}^{T-1} \frac{(T+1)!}{k!} \\ &= \sum_{k=0}^{T} \frac{(T+1)!}{k!}. \end{aligned}$$

□

Now let us consider a particular action sequence **a** in this problem. We have the sequence that always continues, and all others involve continuing a certain number of times, then stopping. Let's analyze how many deviation rules are necessary to check that the sequence that always continues is truly dominated. To do that, we prove the following:

Proposition 2. *In the stopping problem, let* **a** *be the sequence that always continues. For each t, there is only one deviation rule (other than the identity) that removes the sequence $(a_1, \ldots, a_t)$.*

Proof. By definition, if the deviation rule removes $(a_1, \ldots, a_t)$, then it must be absorbing for every **b** with $(b_1, \ldots, b_t) \neq (a_1, \ldots, a_t)$. Thus any sequence that stops before period t must be unaffected by the deviation rule. This also means that any sequence that starts with $(a_1, \ldots, a_t)$ cannot be taken to a sequence that stops before period t, otherwise adaptedness would mean there is a **b** that contradicts our previous statement.

We must also have that **b** is repulsive whenever $(b_1, \ldots, b_t) = (a_1, \ldots, a_t)$, meaning that a sequence that does not stop in the first t periods has to be taken to the sequence that stops exactly at period t. Thus, we are left with only one deviation rule that removes $(a_1, \ldots, a_t)$. □

This immediately implies that for each period t, we have only one deviation rule to consider, totaling T deviation rules to consider.

6 Another potential approach: Backward induction

It is tempting to frame the solution to our problem in a recursive or inductive form. Here we show that a natural backward inductive approach can be useful in thinking about true dominance, and that eventually our notion of deviation rules subsumes the inductive construction.

The following simple result shows how we can derive conclusions about a decision problem by looking at particular subproblems.

Proposition 3 (Informal)**.** *If* $(a_{t+1}, \ldots, a_T)$ *is truly dominated in the subproblem obtained by fixing* $(a_1, \ldots, a_t)$*, then* $(a_1, \ldots, a_T)$ *is truly dominated in the original problem.*

Proposition 3 gives a method of finding truly dominated sequences by backward induction. We first fix $(a_1, \ldots, a_{T-1})$ and then find which actions a_T are truly dominated in the single-period problem that follows.[9] Let $\tilde{A}_T$ be the last period actions that survived (that is, can be rationalized), and now fixing $(a_1, \ldots, a_{T-2})$ we find which sequences $(a_{T-1}, a_T) \in A_{T-1} \times \tilde{A}_T$ are truly dominated in this two-period problem, and so on. This exercise helps the analyst in two ways. First, it directly simplifies her search for the set of action sequences that cannot be rationalized, and second, as we will prove in the next subsection, it informs her that the construction of deviation rules for other action sequences should not take these truly dominated action sequences in their support. For example, if action sequence if the (a_{T-1}, a_T) is truly dominated in the subproblem, the analyst immediately knows that whole action sequence $\mathbf{a}$ is truly dominated in the original problem, and moreover, that in order to construct a deviation rule for any other action sequence $\mathbf{b}$ that may be truly dominated, the associated deviation rule does not have to put any weight on $\mathbf{a}$.

There are, however, two caveats to making backward induction the primary approach in solving our problem. First, in its final stages the method described above can be almost as complex as the original problem. Second, a naive application of it might lead to mis-identification of the set of action sequences that can be rationalized, as the following conjecture exposits:

Conjecture 1. *Suppose that (i) the sequence of actions* $(a_1, \ldots, a_t)$ *can be chosen with positive probability, and (ii)* $(a_{t+1}, \ldots, a_T)$ *can be rationalized in the subproblem obtained by fixing* $(a_1, \ldots, a_t)$*. Then,* $(a_1, \ldots, a_T)$ *can be rationalized.*

This conjecture is false; the decision tree in Figure 4 is a simple counterexample. Both L and R can be chosen in the first period, and in the decision following action L, both l and r can be chosen. This could lead the analyst to believe that (L, r) can be rationalized. However, the deviation rule depicted in the figure shows otherwise. The problem with naive inductive reasoning is that a choice of L can only be rationalized if the agent is sure about the state being the first one; choosing r would then require an inconsistent belief. Such indifference in payoffs

[9] Since this is a "static" problem, it is the same as looking for actions which are strictly dominated by some other action.

(both (L, l) and (R, r) yield 3 in the first state), requires a comparison of the full sequence of actions. Thus, in general, it is apt to define true dominance along the entire sequence of actions and employ the induction argument to construct simple deviation rules whenever possible.

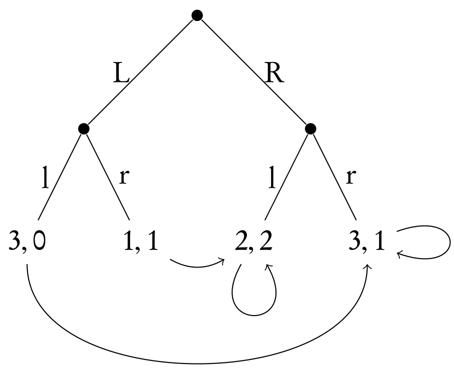


Figure 4: Counter example to naive reasoning in Conjecture 1—the deviation rule shown is adapted and improves upon Lr

7 Final remarks

We conjecture that further improvements can be made in ruling out bad candidate deviation rules and reducing the set to search over even more dramatically. Note, for instance, that nowhere in the description of the problem in Section 4 or in the statement of Theorem 2 did we use payoffs; only the structure of the decision tree has been invoked. It is clear that for specific trees with systematic payoff structures, more precise statements can be made about reducing the complexity of the search for potentially binding deviation rules.

In de Oliveira and Lamba [2025], analogous to Theorem 1 above, we also provide a similar characterization result for rationalizing distributions over action sequences, and further describe how the search for deviation rules can be reduced to a simple linear program. A similar reduction in the set of potentially deviation rules (as Theorem 2 above) for the case of rationalizing distributions., is an interesting open question.

8 Proofs

8.0.1 Simple deviations

We will need a few lemmata before proving Proposition 2. The proof is by induction, and the first lemma is a version of Proposition 2 involving only the first period action.

Lemma 1. *If* $\mathbf{a} = (a_1, \ldots, a_T)$ *is a truly dominated action sequence, then there exists a deviation rule* D *that dominates* $\mathbf{a}$*, satisfying* $D(\mathbf{b}) = I(\mathbf{b})$ *whenever* $b_1 \neq a_1$.

Proof. The argument here is similar to the proof of Proposition 3. Let F be a deviation rule that dominates $\mathbf{a}$ and let

$$D(\mathbf{b}) = \begin{cases} F(\mathbf{b}) & \text{when } b_1 = a_1 \\ I(\mathbf{b}) & \text{when } b_1 \neq a_1 \end{cases}.$$

It is easy to check that D is also adapted and dominates $\mathbf{a}$. To show adaptedness, we can show, as in Proposition 3, that

$$\sum_{c_2,\ldots,c_T} D(\mathbf{c}|\mathbf{b}) = \begin{cases} \sum_{c_2,\ldots,c_T} F(\mathbf{c}|\mathbf{b}) & \text{if } b_1 = a_1 \\ \sum_{c_2,\ldots,c_T} I(\mathbf{c}|\mathbf{b}) & \text{if } b_1 \neq a_1 \end{cases}$$

so adaptedness of D follows from adaptedness of F and I. To see that D dominates $\mathbf{a}$, notice that

$$u(D(\mathbf{b}), \omega) = \begin{cases} u(F(\mathbf{b}), \omega) & \text{when } b_1 = a_1 \\ u(\mathbf{b}, \omega) & \text{when } b_1 \neq a_1 \end{cases}.$$

□

The key idea behind the lemmata that follow is to regard a deviation rule as the transition probabilities of a Markov chain, with the set of action sequences A as the set of “markov states" (not to be confused with the “states of nature", $\omega \in \Omega$). We can then label action sequences according to their properties as markov states:

Definition 7. *Let* D *be a deviation rule. We say that an action sequence* $\mathbf{a} \in A$ *is*

1. recurrent *if, starting from* $\mathbf{a}$*, the probability of eventually returning to* $\mathbf{a}$ *is one;*
2. transient *if, starting from* $\mathbf{a}$*, the probability of eventually returning to* $\mathbf{a}$ *is less than one;*
3. absorbing *if* $D(\mathbf{a}|\mathbf{a}) = 1$*;*
4. repulsive *if, for every* $\mathbf{b} \in A$, $D(\mathbf{a}|\mathbf{b}) = 0$.

Note that absorbing implies recurrent and repulsive implies transient, but the converse in each case is not true. This is standard terminology for Markov chains, with the exception of the definition of “repulsive" (see Kemeny, Snell, and Knapp [1976]). The idea that follows relates payoff dominance of action sequences to their properties as Markov states. The intuition here is that if D dominates $\mathbf{a}$ then it must have a tendency to move the agent away from $\mathbf{a}$, so that $\mathbf{a}$ will be transient.

Lemma 2. *Let D be a deviation rule that dominates* $\mathbf{a}$*. Then* $\mathbf{a}$ *is a transient state for D.*

Proof. If $\mathbf{a}$ is a recurrent state, then there exists a stationary probability $\alpha \in \Delta(A)$ such that $D \circ \alpha = \alpha$ and $\alpha(\mathbf{a}) > 0$ (Theorem 6.9 in Kemeny, Snell, and Knapp [1976]). Now define a distribution $\gamma \in \Delta(A \times \Omega)$ by $\gamma(\mathbf{b}, \omega) = \alpha(\mathbf{b})\, p(\omega)$, where $p \in \Delta(\Omega)$ is arbitrary. Then $\gamma(\mathbf{a}) > 0$ and $\mathbb{E}_\gamma\left[u\left(D(\mathbf{b}), \omega\right)\right] = \mathbb{E}_\gamma\left[u(\mathbf{b}, \omega)\right]$, which contradicts the fact that D dominates $\mathbf{a}$. □

The following fundamental construction is what allows us to turn transient action sequences to the stronger property of being repulsive. Here we use the notation D^k to mean the composition of D with itself k times.

Lemma 3. *Let D be any deviation rule. Then*

$$D^\infty(\mathbf{c}|\mathbf{b}) = \lim_n \frac{1}{n} \sum_{k=1}^{n} D^k(\mathbf{c}|\mathbf{b})$$

is a well defined deviation rule. Moreover,

1. $D^\infty = D \circ D^\infty = D^\infty \circ D = D^\infty \circ D^\infty$
2. *If* $\mathbf{b}$ *is dominated by D, then it is also dominated by D^∞;*
3. *If* $\mathbf{b}$ *is transient for D, then it is repulsive for D^∞;*
4. *If* $\mathbf{b}$ *is absorbing for D, then it is also absorbing for D^∞*

Proof. The proof that the limit exists and of (1) follows that of the Ergodic Theorem for Markov Chains, (see Theorem 6.1 in Kemeny, Snell, and Knapp [1976]). The proof of (4) follows straightforwardly, by construction. Here we prove parts (2) and (3).

If D dominates $\mathbf{a}$, then for every $\mathbf{b} \in A$ and $\omega \in \Omega$,

$$\begin{aligned} u\left(D^k(\mathbf{b}), \omega\right) &= \sum_{\mathbf{c}} u\left(D(\mathbf{c}), \omega\right) D^{k-1}(\mathbf{c}|\mathbf{b}) \\ &\geqslant \sum_{\mathbf{c}} u(\mathbf{c}, \omega)\, D^{k-1}(\mathbf{c}|\mathbf{b}) \\ &\vdots \\ &\geqslant u\left(D(\mathbf{b}), \omega\right) \end{aligned}$$

This shows that D^k dominates $\mathbf{a}$ as well. Combining the inequalities for different k and taking the limit, we conclude that $u\left(D^{\infty}\left(\mathbf{b}\right),\omega\right) \geqslant u\left(D\left(\mathbf{b}\right),\omega\right)$ as well, so D^{∞} dominates $\mathbf{a}$.

To prove (3), suppose $\mathbf{b}$ is transient. We must show that $D^{\infty}(\mathbf{b}|\mathbf{a}) = 0$ for all $\mathbf{a} \in A$. If $\mathbf{a}$ is recurrent, then we must have $D^k\left(\mathbf{b}|\mathbf{a}\right) = 0$ for all k, otherwise $\mathbf{b}$ would be recurrent as well (see Lemma 4.23 in Kemeny, Snell, and Knapp [1976]). If $\mathbf{a}$ is transient, then we have that $\lim_k D^k\left(\mathbf{b}|\mathbf{a}\right) = 0$ (by Proposition 5.3 in Kemeny, Snell, and Knapp [1976]). Hence, in either case, we have that $D^{\infty}\left(\mathbf{b}|\mathbf{a}\right) = 0$, so $\mathbf{b}$ is repulsive for D^{∞}. □

A given deviation rule D induces a deviation rule up to each period t, which we will denote by D_t. That is,

$$D_t\left(c_1,\ldots,c_t|b_1,\ldots,b_t\right) = \sum_{c_{t+1},\ldots,c_T} D\left(c_1,\ldots,c_T|b_1,\ldots,b_T\right)$$

where $b_{t+1},\ldots,b_T$ can be chosen arbitrarily, since D is adapted.

Lemma 4. *If* $D\left(\mathbf{b}\right) = I\left(\mathbf{b}\right)$ *whenever* $b_1 \neq a_1$ *and* $D_1\left(a_1|a_1\right) < 1$, *then* $D_1^{\infty}\left(a_1|a_1\right) = 0$.

Proof. Let $\lambda = D_1\left(a_1|a_1\right) < 1$ and notice that

$$D_1^2\left(a_1|a_1\right) = \sum_{b_1} D_1\left(a_1|b_1\right) D_1\left(b_1|a_1\right) = D_1\left(a_1|a_1\right) D_1\left(a_1|a_1\right)$$

since $D_1\left(a_1|b_1\right) = 0$ whenever $b_1 \neq a_1$. From this, we deduce that $D_1^k\left(a_1|a_1\right) = \lambda^k$. Hence

$$D_1^{\infty}\left(a_1|a_1\right) = \lim_n \frac{1}{n}\sum_{k=0}^{n-1} D_1^k\left(a_1|a_1\right) = \lim_n \frac{1}{n}\left(\frac{1-\lambda^n}{1-\lambda}\right) = 0.$$

□

Proof of Proposition 2. Suppose $\mathbf{a} = (a_1,\ldots,a_T)$ is truly dominated. We want to show that there exists a deviation rule D that dominates $\mathbf{a}$ and there exists a t such that D removes $(a_1,\ldots,a_t)$. We proceed by induction on T. The idea is that starting from a deviation rule D that dominates $\mathbf{a}$, we will show that D^{∞} removes $(a_1,\ldots,a_t)$ for some t.

Let $T = 1$. Then, obviously there exists a deviation rule D such that $D\left(b_1\right) = I\left(b_1\right)$ whenever $b_1 \neq a_1$ and $D\left(a_1|a_1\right) < 1$. Thus from Lemma 4, we have $D_1^{\infty}\left(a_1|a_1\right) = 0$.

Next, suppose that result holds for $T-1$ where $T \geq 2$. We want to show that it is true for T. Using Lemma 1, let D dominate $\mathbf{a}$ such that $D\left(\mathbf{b}\right) = I\left(\mathbf{b}\right)$ whenever $b_1 \neq a_1$. Now there are two possible cases to consider: $D_1\left(a_1|a_1\right) < 1$ and $D_1\left(a_1|a_1\right) = 1$.

If $D_1(a_1|a_1) < 1$, then from Lemma 4 we know that $D_1^\infty(a_1|a_1) = 0$, and thus $D^\infty(\mathbf{b}|\mathbf{a}) = 0$ for $b_1 = a_1$. Moreover, recollect from Lemma 3 part (4), we know that $D^\infty(\mathbf{b}) = I(\mathbf{b})$ for $b_1 \neq a_1$. Therefore, we can conclude that D^∞ removes a_1.

Now, suppose $D_1(a_1|a_1) = 1$, which means that the deviation rule D takes every sequence starting with a_1 to another sequence starting with a_1. Therefore, it naturally defines a deviation rule for the subproblem that fixes a_1. Further, it is easy to see that the induced deviation rule shows $(a_2, \ldots, a_T)$ to be truly dominated in the subproblem (just reverse the construction in the proof of Proposition 3).

Note that the subproblem is of length $T-1$, so using the induction hypothesis, we know that there exists a deviation rule F and a t such that F dominates $(a_2, \ldots, a_T)$ and F removes $(a_2, \ldots, a_t)$. Thus, in original problem the deviation rule G, defined by $G(\mathbf{b}) = I(\mathbf{b})$ whenever $b_1 \neq a_1$, and $G(\mathbf{c}|\mathbf{b}) = F(c_2, \ldots, c_T|b_2, \ldots, b_T)$ if $b_1 = a_1$, dominates $\mathbf{a}$ and removes $(a_1, \ldots, a_t)$. □

8.0.2 Backward induction

When stating Proposition 3, we informally referred to a "subproblem". We begin by defining this concept precisely.

Definition 8. *We refer to the collection of action sets $A_1, \ldots, A_T$, together with the utility function $u : A \times \Omega \to \mathbb{R}$, as the agent's* decision problem. *The* subproblem *obtained by fixing $(a_1, \ldots, a_t)$ is the subcollection of action sets $A_{t+1}, \ldots, A_T$, together with the utility function $v : A_{t+1} \times \cdots \times A_T \times \Omega \to \mathbb{R}$ defined by*

$$v(a_{t+1}, \ldots, a_T, \omega) = u(a_1, \ldots, a_t, a_{t+1}, \ldots, a_T, \omega).$$

Thus, a sequence $(a_{t+1}, \ldots, a_T)$ is truly dominated in the subproblem if there exists a deviation rule $D : A_{t+1} \times \cdots \times A_T \to \Delta(A_{t+1} \times \cdots \times A_T)$ such that

$$\begin{aligned} v(D(a_{t+1}, \ldots, a_T), \omega) &> v(a_{t+1}, \ldots, a_T, \omega) \quad \text{for all } \omega \in \Omega \text{ and} \\ v(D(b_{t+1}, \ldots, b_T), \omega) &\geqslant v(b_{t+1}, \ldots, b_T, \omega) \quad \text{for all } b_{t+1} \in A_{t+1}, \ldots, b_T \in A_T, \omega \in \Omega. \end{aligned}$$

Proof of Proposition 3 . Here we will use the following notation: given $\mathbf{b} = (b_1, \ldots, b_T)$, we will let $\mathbf{b}\big|_t = (b_1, \ldots, b_t)$.

Let $D : A_{t+1} \times \cdots \times A_T \to \Delta(A_{t+1} \times \cdots \times A_T)$ be a deviation rule for the subproblem such that D dominates $(a_{t+1}, \ldots, a_T)$. Then we can define $F : A \to \Delta(A)$ to recommend the same deviations as D in the subproblem, and recommend no deviations elsewhere. Formally,

$$F(\mathbf{c}|\mathbf{b}) = \begin{cases} D(c_{t+1}, \ldots, c_T | b_{t+1}, \ldots, b_T) & \text{if } \mathbf{b}\big|_t = \mathbf{a}|_t = \mathbf{c}|_t \\ I(\mathbf{c}|\mathbf{b}) & \text{otherwise} \end{cases}$$

where I is the identity (= 1 if $\mathbf{c} = \mathbf{b}$, and zero otherwise). Also, recall that for a stochastic mapping, $F(\mathbf{c}|\mathbf{b})$ is the probability that $F(\mathbf{b}) \in \Delta(A)$ puts on the action sequence $\mathbf{c}$. Now, we claim that F is adapted and also dominates $\mathbf{a}$. To show that F is adapted, we must show that $\sum_{c_{s+1},\ldots,c_T} F(\mathbf{c}|\mathbf{b})$ does not depend on $(b_{s+1}, \ldots, b_T)$. We show this separately for $s \geqslant t$ and $s < t$.

So fix $\mathbf{b} = (b_1, \ldots, b_T)$ and $\mathbf{c}|_s = (c_1, \ldots, c_s)$ and suppose $s \geqslant t$. Then $\mathbf{b}\big|_t$ and $\mathbf{c}|_t$ are uniquely determined, and every term in the sum for F is given by D or every term in the sum is given by I. Hence we have

$$\sum_{c_{s+1},\ldots,c_T} F(\mathbf{c}|\mathbf{b}) = \begin{cases} \sum_{c_{s+1},\ldots,c_T} D(c_{t+1}, \ldots, c_T | b_{t+1}, \ldots, b_T) & \text{if } \mathbf{b}\big|_t = \mathbf{a}|_t = \mathbf{c}|_t \\ \sum_{c_{s+1},\ldots,c_T} I(\mathbf{c}|\mathbf{b}) & \text{otherwise} \end{cases}$$

In either of those cases, the sum on the right does not depend on $(b_{s+1}, \ldots, b_T)$, since both D and I are adapted.

Now suppose $s < t$. If $\mathbf{c}|_s \neq \mathbf{a}|_s$ or $\mathbf{b}\big|_t \neq \mathbf{a}|_t$, then we already know that $F(\mathbf{c}|\mathbf{b}) = I(\mathbf{c}|\mathbf{b})$, so again $\sum_{c_{s+1},\ldots,c_T} F(\mathbf{c}|\mathbf{b}) = \sum_{c_{s+1},\ldots,c_T} I(\mathbf{c}|\mathbf{b})$ does not depend on $(b_{s+1}, \ldots, b_T)$. But if $\mathbf{c}|_s = \mathbf{a}|_s$ and $\mathbf{b}\big|_t = \mathbf{a}|_t$, some terms of the sum have $\mathbf{c}|_t = \mathbf{a}|_t$ and others have $\mathbf{c}|_t \neq \mathbf{a}|_t$. In that case, we can write

$$\begin{aligned} \sum_{c_{s+1},\ldots,c_T} F(\mathbf{c}|\mathbf{b}) &= \sum_{\{\mathbf{c}:\mathbf{c}|_t = \mathbf{a}|_t\}} D(c_{t+1}, \ldots, c_T | b_{t+1}, \ldots, b_T) + \sum_{\{\mathbf{c}:\mathbf{c}|_t \neq \mathbf{a}|_t\}} I(\mathbf{c}|\mathbf{b}) \\ &= \sum_{c_{t+1},\ldots,c_T} D(c_{t+1}, \ldots, c_T | b_{t+1}, \ldots, b_T) + \sum_{\{\mathbf{c}|\mathbf{c}|_t \neq \mathbf{a}|_t\}} I(\mathbf{c}|\mathbf{a}) \\ &= 1 + 0 \\ &= \sum_{c_{s+1},\ldots,c_T} I(\mathbf{c}|\mathbf{b}). \end{aligned}$$

For the third equality, the first term is 1 because we are summing over the entire support of that distribution, and the second term is zero because when $\mathbf{c}|_t \neq \mathbf{a}|_t$ the identity map gives a value of zero. So we have shown that, whenever $s < t$, we have $\sum_{c_{s+1},\ldots,c_T} F(\mathbf{c}|\mathbf{b}) = \sum_{c_{s+1},\ldots,c_T} I(\mathbf{c}|\mathbf{b})$.

Summarizing all cases, we have shown

$$\sum_{c_{s+1},\ldots,c_T} F(\mathbf{c}|\mathbf{b}) = \begin{cases} \sum_{c_{s+1},\ldots,c_T} D(c_{t+1},\ldots,c_T|b_{t+1},\ldots,b_T) & \text{if } s \geqslant t \text{ and } \mathbf{b}\big|_t = \mathbf{a}|_t = \mathbf{c}|_t \\ \sum_{c_{s+1},\ldots,c_T} I(\mathbf{c}|\mathbf{b}) & \text{otherwise.} \end{cases}$$

In either case, the sum on the right-hand side does not depend on $(b_{s+1},\ldots,b_T)$, since D and I are adapted. Hence, F is adapted as well.

To see that F dominates **a**, notice that $u(F(\mathbf{b}),\omega) = u(I(\mathbf{b}),\omega) = u(\mathbf{b},\omega)$ if $\mathbf{b}\big|_t \neq \mathbf{a}|_t$. When $\mathbf{b}\big|_t = \mathbf{a}|_t$, we have that $u(F(\mathbf{b}),\omega) \geqslant u(\mathbf{b},\omega)$ for all **b** and with strict inequality for $\mathbf{b} = \mathbf{a}$, since there F recommends the same deviation as D. □

References

H. de Oliveira. Blackwell's informativeness theorem using diagrams. *Games and Economic Behavior*, 109:126–131, 2018.

H. de Oliveira and R. Lamba. Rationalizing dynamic choices. Fundação Getúlio Vargas-EESP and Cornell University, 2025.

J. G. Kemeny, J. L. Snell, and A. W. Knapp. *Denumerable Markov chains*. Springer-Verlag, 1976.