How to Buy Advice*

Ronen Gradwohl Yuval Salant †

First version: January 3, 2011
This version: September 20, 2011

Abstract

A decision maker, whose payoff is influenced by an unknown stochastic process, seeks the advice of an advisor, who may be informed about the process. We identify a sufficient condition on the correlation between the advisor's information and the true stochastic process, called conservativeness, for which there exists a strategy D of the decision maker that will yield him an almost first-best payoff in every period. We also demonstrate that without conservativeness no strategy can approximate the first-best payoff.

The belief-free strategy D satisfies various desirable properties. It only requires a fixed budget – regardless of the realizations of the stochastic process and whether or not the advisor is actually informed about it, the total payoff to the decision maker will never fall below a fixed threshold. Moreover, per-period compensation to the advisor is independent of the present realization of the process, and depends solely on the expected value of the advice as reported by the advisor.

*We thank Ehud Kalai, Wojciech Olszewski, and Rakesh Vohra for fruitful discussions.

†Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. E-mails: r-gradwohl@kellogg.northwestern.edu and y-salant@kellogg.northwestern.edu.

1 Introduction

A decision maker (DM) faces uncertainty about events that unfold over time and influence his payoff. He seeks the advice of an advisor who claims to be informed about the process that governs the events. In actuality the advisor may either be informed or uninformed, and the DM is uncertain about this as well. When interacting with the advisor, can the DM obtain a "high" payoff in case the advisor is informed while simultaneously not losing "too much" in case the advisor is uninformed?

This question arises, for example, when an investor considers hiring a financial advisor to manage his portfolio, or when a firm considers hiring a consultant to get advice on certain aspects of its business. In both situations, the advisor may be informed or uninformed about the process that influences the payoff of the DM, and the DM may be uncertain about that as well as about the underlying process.[1] In both situations, it is hard to evaluate the quality of the advice without actually following it, yet following it may be costly to the DM. Finally, in both situations, the interaction with the advisor is potentially repeated over time, where in every period the advisor provides some advice and gets compensated for it.

We study a repeated interaction between a DM and an advisor with all the above features. Our first observation is that even when the advisor is informed, it may be impossible to achieve a high payoff in the interaction with him without the risk of losing too much. Consider the following example:

Example 1.1 Nature is a stochastic process that realizes in every period one of two states, H or L. Up to period T, the state is realized according to a coin toss that assigns probability $\frac{3}{4}$ to state H and $\frac{1}{4}$ to state L. From period T onward, the process becomes a deterministic sequence of H's and L's.

An advisor believes that Nature is a deterministic sequence of H's and L's: up to period T he believes the state is always H, and thus he correctly estimates the direction of the bias but overestimates its magnitude. From period $T+1$, he accurately predicts Nature's state.

A DM is uncertain about all of the above and his uncertainty cannot be captured by a prior. In every state, he may bet on the state or stay out. If he bets, he wins \$1 if correct and -\$3 if incorrect. If he stays out, he receives a sure payoff of \$0 and has no indication of the state realization. The DM has a "budget" of \$$B$ that reflects the

[1] Such uncertainty may be alleviated in environments where developing reputation for being informed is possible.

maximum amount the he is willing to lose in the interaction with the advisor.

Let us verify that even though the advisor above is informed, it is impossible for a DM to identify the period T from which the advisor accurately predicts the state without exceeding his budget \$$B$.

Suppose, for simplicity, that the advisor charges a fixed fee of \$$c$ per period for his advice (in case the DM decides to follow it) and that the DM uses a deterministic strategy in the interaction with the advisor.[2] Denote by $T_{\max}$ the smallest integer that satisfies that the DM consulted with the advisor and invested $B/c+1$ times up to period $T_{\max}$. Clearly, $T_{\max}$ exists and is finite or else the DM will not be able to identify T and obtain a high payoff, if T were large enough. But if $T_{\max} < T$ then in expectation the DM will exceed his budget by period $T_{\max}$, because every time he consults with the advisor and invests he obtains an expected payoff of \$0 and has to pay the advisor \$$c$. $\diamondsuit$

The advisor in Example 1.1 is informed: Up to period T, he identifies correctly the direction of the coin's bias but overestimates it; from period $T+1$ he accurately predicts the state. To achieve a high payoff, the DM needs to identify the period T, and to do so, the DM must consult with the advisor every once in a while. The advisor overestimates the value of his advice (he believes it is 1 while it is actually 0) and recommends that the DM invest. But every time the DM invests prior to period T he loses money, and thus may use up his budget. This is a general observation: To achieve a high payoff the DM needs to consult with the advisor every once in a while, but if the advisor overestimates the true value of his advice in these periods and recommends that the DM invest this may harm the DM.

We therefore focus on advisors who underestimate the value of their advice or only "slightly" overestimate it. We say that an advisor is *conservative* if the advisor's assessed value of his best advice (according to the advisor's information) does not overestimate the true expected value of the same advice (according to Nature's process) by too much.

Conservativeness comes up naturally in the context of learning.

Example 1.2 In every period, Nature realizes either the state H or the state L according to an i.i.d. coin toss. The advisor knows that Nature's process is i.i.d., and

[2] In Section 3.2 we establish an impossibility result for any strategy of the DM and any limited-liability compensation scheme.

he has some prior over the bias of the coin that puts non-zero weight on the true bias. He observes some realizations of the coin prior to his interaction with the DM.

Similarly to Example 1.1, the DM is uncertain about all of the above and his uncertainty cannot be captured by a prior. He may bet on the state (and get a positive payoff if correct) or stay out.

Then, with high probability, the advisor will be able to approximate the true bias of the coin prior to interacting with the DM. His assessed value of his advice will approximate the true value of his advice according to Nature's process. Hence, he is conservative. $\diamondsuit$

Conservativeness also arises when the advisor is more optimistic than Nature about the likelihood of outcomes with negative payoffs and is more pessimistic than Nature about the likelihood of outcomes with positive payoff, given his best advice. For example,

Example 1.3 Nature is a stochastic process that realizes in every period one of two states, H or L, according to a coin toss that assigns probability $\frac{3}{4}$ to state H and $\frac{1}{4}$ to state L. The advisor believes that the bias to state H is actually $\frac{2}{3}$.

The DM may bet on the state or stay out. If he bets, he wins \$1 if correct and -\$1 if incorrect. If he stays out, he receives a sure payoff of 0.

The advisor's best advice is to bet on H. He underestimates the probability of the outcome with positive payoff: He assigns a probability of $\frac{2}{3}$ to the positive payoff as compared to the true probability of $\frac{3}{4}$ and overestimates the outcome with the negative payoff. Hence, he is conservative. $\diamondsuit$

Our main result answers our initial question affirmatively – the DM *can* obtain a "high" payoff in case the advisor is informed while simultaneously not losing "too much" in case the advisor is uninformed – when informed advisors are ones that are conservative. Specifically, there exists a simple belief-free calibration strategy D of the DM and a zero-liability compensation scheme to the advisor that achieve the following:

(1) If the advisor is truthful (i.e., never distorts his view of the underlying process) and conservative, then after an initial test period, the sum of the DM's per-period payoffs *up to every period* is *close with high probability* to the sum of payoffs the DM would have expected to obtain if he knew the advisor's process;

(2) If the advisor is conservative and strategic (i.e., provides advice that improves his own payoff with high probability over truthfulness), then the DM's payoff guarantee can only increase;

(3) The budget spent by the DM never exceeds a fixed and bounded amount that depends on how close to the optimal payoff the DM wants to be and on the probability with which he wants to obtain it. Importantly, this budget is independent of the number of periods t.

Put differently, there exists a strategy that enables the DM to approximate the sum of payoffs he could have achieved if he knew the information of the advisor and used it optimally in case the advisor is conservative, and to lose up to a fixed and bounded amount otherwise. This is possible even though the only instruments available to the DM in interacting with the advisor are a zero-liability compensation scheme and the ability to stop interacting with him.

We also extend Example 1.1 and show that the assumption of conservativeness is necessary in the sense that for any limited-liability compensation scheme and strategy of the DM, there exist an advisor that is "far" from being conservative (yet is informed in the sense of Example 1.1) and a process of Nature for which the DM will not even approximate the first-best.

The strategy D that achieves the guarantees of the positive result above has two phases: In the first "test" phase, the DM buys information from the advisor whenever the advisor claims the value of the information is larger than some threshold value. In the second "calibration" phase, the DM buys information from the advisor whenever the advisor claims its value is larger than that threshold, and as long as the past recommendations of the advisor are close to the payoff realizations of the DM. Note that D does not require the DM to form beliefs about the true underlying process, the advisor's process, the correlation between the two, or the advisor's strategy, which may be a challenging task in complex environments.

The corresponding compensation scheme is to award the advisor a small fraction of what he claims is the expected value of his advice in case the DM decides to follow it, and zero otherwise. This compensation scheme has zero liability – payments are made from the DM to the advisor, and never vice versa.[3] In addition, compensation is not outcome-based in the sense that it does not depend on whether the advice turns out to be valuable, but only on the expected value of the advice as stated by the advisor. These features are reminiscent of portfolio management contracts, in

[3] See Chassang (2011) for further discussion of limited-liability contracts.

which a manager's fee depends only on the amount invested in each period and not on the realized returns from that period (see for example Farnsworth, 2011).

We proceed as follows. Related literature is surveyed below. Section 2 presents the model. Because our goal is to establish a possibility result, we fix the compensation scheme to be the one mentioned above. Section 3 presents our main results. Section 4 comments on possible extensions of our analysis.

1.1 Related literature

Our results are related to the recent literature on testing experts. In this literature, a DM desires a test that will determine whether an expert is knowledgable and knows Nature's stochastic process, or whether he is ignorant and does not know Nature's process. Foster and Vohra (1998), Fudenberg and Levine (1999), Lehrer (2001), Sandroni (2003), Sandroni, Smorodinsky, and Vohra (2003), Vovk and Shafer (2005), Olszewski and Sandroni (2008, 2009), and Shmaya (2008) establish in various settings that any test that passes a knowledgable expert is also manipulable: An ignorant expert can strategically generate predictions so that he matches the performance of a knowledgable expert.

In order to overcome this impossibility result, various authors have relaxed some of its underlying assumptions to obtain a non-manipulable test. Among the relevant papers are Dekel and Feinberg (2006), Olszewski and Sandroni (2008, 2009), Al-Najjar and Weinstein (2008), Al-Najjar et al. (2010), Fortnow and Vohra (2009), Hu and Shmaya (2010), Echenique and Shmaya (2008), and Olszewski and Pęski (forthcoming). However, while the distinction between an ignorant expert and one who is completely informed about the true process makes impossibility results stronger, this extreme distinction makes the possibility results rather weak. In particular, the non-manipulable test is guaranteed to pass only *completely* knowledgeable experts, and may fail experts who are *almost* fully knowledgeable (such as the conservative advisors in this paper).

Of the above papers, Echenique and Shmaya (2008) and Olszewski and Pęski (forthcoming) model explicitly how the expert's advice influences the payoff of the DM. Both papers establish that there exists a test that passes a knowledgable expert, and if it also passes an ignorant expert then the predictions of that ignorant expert do not harm the DM too much.

In Echenique and Shmaya's model (2008), a DM has a theory π about how certain events will unfold over time. The DM needs to decide whether to replace that theory with a new theory ν offered by an expert, as he will then use the selected theory to make payoff-relevant choices. Echenique and Shmaya show that there exists a test that guarantees the DM the following: (1) The test passes ν with certainty if ν is a true theory, and (2) if the test passes some theory ν, then an infinitely patient DM who behaves according to ν will obtain an expected payoff that weakly exceeds his expected payoff under π (where the expectation is taken with respect to the DM's original theory π).

Olszewski and Pęski's (forthcoming) principal-agent model is more closely related to our model. In their model, the DM needs to take an action in each period, and his per-period payoff depends on his action and an unknown state. The DM seeks the advice of an expert who may be knowledgable or ignorant. He offers the potential expert a menu of contracts, each defining the periods in which the expert will be required to provide predictions as well as the expert's compensation, a function of his predictions and the realized outcomes. The expert chooses a contract from the menu, and then provides predictions and gets compensated accordingly. Olszewski and Pęski establish that if the DM evaluates infinite payoff sequences according to their limit average, then there exists a menu of contracts that enables a DM to achieve the following: (1) If the expert is knowledgable the DM obtains a payoff close to the payoff he would obtain if he knew Nature's process, and (2) if the expert is ignorant the DM's payoff can fall only marginally below his outside option.

We establish stronger possibility results. First, our results apply not only when the advisor is fully informed, but also when he is partially informed and conservative. Second, the advisor in our model only needs to form beliefs about Nature's process in the *next* period, whereas in Olszewski and Pęski (forthcoming), the advisor needs to form beliefs about Nature's *entire* process in order to choose a contract. Third, we show that there exists a strategy that enables the DM, after an initial test period, to extract almost full surplus *up to every period* in the interaction rather than only in the limit. Fourth, this strategy uses a budget that is bounded and fixed (while in Olszewski and Pęski, the budget required for implementing any contract in the menu is not bounded even though its limit average is negligible). Finally, compensation to the advisor in our model does not depend on the realized state of nature, which provides stronger incentives to the advisor to mis-report his information.

Our analysis is also related to the famous theorem of Hannan (1957). Hannan's theorem states that if a DM receives recommendations from a fixed number of experts, then he has a decision scheme that approximates the payoff he would receive from following the recommendations of the "best" expert among them. The approximation is such that for any number of periods t, the DM's payoff is the payoff from the best advice minus $O(\sqrt{t})$. There are many variations on this theorem, many of which are surveyed by Cesa-Bianchi and Lugosi (2006). The variation that is most relevant to the current work is that of Auer et al. (2002), in which the DM must choose exactly one expert from whom to receive a recommendation in each round. He does not see the recommendations of non-chosen experts in that round. This corresponds to our model in the sense that if the DM does not purchase advice in a certain round, then he does not observe the realization of Nature in that round.

One can embed our problem into this framework by constructing two experts: one expert recommends the DM to purchase advice from the advisor whenever the declared value of the advice is positive. The other expert always recommends to stay out. The work of Auer et al. (2002) implies that there exists a strategy for the DM in which he always obtains at least the value of the advisor minus $O(\sqrt{t})$ or, if that value is negative, at least the value of staying out (zero) minus $O(\sqrt{t})$.

There are two main differences between our positive results and those of Auer et al. (2002). First, Auer et al. (2002) do not entertain the possibility of strategic experts that may report untruthfully. Second, the budget needed to implement the Auer et al. (2002) strategy is unbounded (i.e., the required budget is $O(\sqrt{t})$), as opposed to our result, in which the DM needs only some *fixed* budget.

2 Model

Nature. Nature is a stochastic process $\mathbf{N} = (\mathbf{N}_1, \mathbf{N}_2, \ldots)$ of random variables with outcomes in a finite set R. Let $r_t \in R$ be the realization of Nature's process in period t, and $r^t = (r_1, \ldots, r_t) \in R^t$ be a vector of realizations in the first t periods. For $t > 1$, the random variable $\mathbf{N}_t = \mathbf{N}_t(r^{t-1})$ may depend on the realizations in periods $1, \ldots, t-1$.

Advisor. An advisor is an agent whose complete view of Nature is captured by a stochastic process $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \ldots)$ of random variables with outcomes in R. The random variable $\mathbf{A}_t = \mathbf{A}_t(r^{t-1})$ may depend on Nature's past realizations, and reflects the advisor's view of $\mathbf{N}_t(r^{t-1})$.

Decision maker. In every period t, the DM decides whether to bet on the outcome of Nature's process in that period or stay out. If he stays out in period t, his payoff is 0 in that period.[4] Otherwise, he chooses a bet z_t from a finite set Z of bets and obtains a per-period payoff of $u(z_t, r_t) \in [-a, b]$, where r_t is the realization of Nature's process in period t. Note that the per-period payoff of the DM is bounded. The DM has no knowledge of Nature's process, the advisor's process, or the correlation between the two. The DM's preference is to stay out without obtaining additional information.

We now describe the interaction between the DM and the advisor. Because our goal is to establish a possibility result, we fix the compensation scheme to the advisor to be the one discussed in the introduction. According to this scheme, whenever the DM decides to follow the advisor's advice, he pays the advisor a small fraction α of the expected per-period net gain, as claimed by the advisor, of following his advice. The fraction α may be an industry standard or the outcome of a bargaining process between the DM and the advisor that we do not model.

Interaction. In every period t, the advisor provides a prediction (v_t, z_t) specifying the maximal expected value v_t of betting in period t according to his information, and the bet z_t that achieves v_t in expectation. If the DM decides not to follow the advice, the period ends and both the DM and the advisor get a payoff of 0. Otherwise, the DM pays the advisor $\alpha \cdot v_t$, observes the realization r_t from $\mathbf{N}_t(r^{t-1})$, and obtains a payoff of $u(z_t, r_t)$. That is, the DM needs to invest in order to observe Nature's realization: This corresponds to the observation that it is often hard to evaluate the quality of advice without actually following it.[5]

Strategies. The advisor's strategy is a sequence of declarations $\{(v_t, z_t)\}_t$, where period t's declaration is a function of (1) the $t-1$ past declarations of the advisor, (2) the $t-1$ past decisions of the DM about whether to buy the information or not,

[4] We extend the analysis to cases in which the outside option is some arbitrary fixed number in section 4.

[5] An alternative setting would be one in which a prediction is a distribution over the possible realizations and in which the DM need not follow the action z_t specified by the advisor. Another alternative setting would be one in which the DM always observes the realization of Nature, but the advisor provides only the expected value v_t of his advice up front. The recommended bet z_t that results in the expected value v_t is communicated only if the DM decides to purchase the information. In both settings our possibility results continue to hold, although it may be possible to obtain a somewhat tighter approximation of the first-best payoff.

and (3) the $t-1$ past realizations of Nature. The DM's strategy is a sequence of binary decisions $\{d_t\}_t$, $d_t \in \{0,1\}$ about whether to use the advisor's information or not, where d_t is a function of (1) the $t-1$ past decisions of the DM, (2) the $t-1$ past declarations of the advisor, and (3) the past realizations in all the periods $j \leq t-1$ in which $d_j = 1$.

Note that DM's strategy is conditioned only on realizations in periods in which he invests. Also note that we do not specify the beliefs of the DM regarding Nature's process, the advisor's process, or the correlation between the two. Our focus will be on identifying a strategy for the DM that performs well independently of such beliefs.

Payoffs. Fix a sequence of realizations r^t, and (pure) strategies $a^t = \{(v_j, z_j)\}_{j \leq t}$ and $d^t = \{d_j\}_{j \leq t}$ of the advisor and the DM respectively.

The advisor's payoff in period j is αv_j if the DM decides to bet according to his advice, and 0 otherwise. Thus, the advisor's payoff up to period t is $p^{A,t}(a^t, d^t, r^t) = \sum_{j=1}^{t} d_j(\alpha v_j)$. When the strategies and realizations are clear from the context, we omit them and simply write $p^{A,t}$.

The DM's payoff in period j if he decides to bet is $u(z_j, r_j) - \alpha v_j$, where z_j is the action recommended by the advisor and r_j is the realization of Nature's process in period j. Otherwise, his payoff is 0. Thus, the DM's payoff to period t is $p^{DM,t} = p^{DM,t}(a^t, d^t, r^t) = \sum_{j=1}^{t} d_j(u(z_j, r_j) - \alpha v_j)$. Note that we are implicitly assuming here that the DM follows the advice he obtains. This only makes our positive result stronger, and for our negative result we dispense with this assumption.

3 Analysis

In this section, we design a strategy for the DM that achieves two goals. First, it obtains a payoff that is close to the first-best payoff in case the advisor is conservative (to be defined below). This requirement is in the spirit of the requirement in the expert-testing literature that a test will pass an expert who knows Nature's process. Second, it bounds the DM's realized loss when interacting with any other advisor. This is similar to the requirement that a test will fail an ignorant expert in the expert-testing literature.

We also show that the assumption of conservativeness is necessary in the sense that for any limited-liability compensation scheme and strategy of the DM, there exist an advisor that is informed in a non-conservative way and a process of Nature for which the DM will not even approximate the first-best.

We begin by defining the notion of first-best payoff.

First-best payoff. In assessing the expected payoff of the DM from betting in some period j, there are at least two issues to consider. First, the expected payoff of the DM is bounded above by the expected payoff of the best bet according to Nature's process, i.e., it is bounded above by $\max_{z \in Z} E(u(z, \mathbf{N}_j(r^{j-1})))$, where $E(u(z, \mathbf{X}))$ denotes the expected payoff of the bet z, taken with respect to the process $\mathbf{X}$. Second, since the only information available to the DM is that of the advisor, the DM cannot expect to get a payoff that is higher than that of the payoff-maximizing bet according to the advisor's process, i.e., $\max_{z \in Z} E(u(z, \mathbf{A}_j(r^{j-1})))$.

We thus define the first-best per-period payoff from betting to be

$$\mathrm{val}_j(r^{j-1}) = \min\left\{\max_{z \in Z} E(u(z, \mathbf{N}_j(r^{j-1}))), \max_{z \in Z} E(u(z, \mathbf{A}_j(r^{j-1})))\right\}.$$

Of course, if that value is negative, the DM can stay out and obtain a payoff of 0. He can thus aim to achieve a payoff of at most $\max\{0, \mathrm{val}_j(r^{j-1})\}$ in every period j. Since the DM also has to compensate the advisor for the advice, we define the first-best payoff up to period t on the history $r^{t-1} = (r_1, \ldots, r_{t-1})$ to be

$$\mathrm{FB}_t(r^{t-1}) = \sum_{j=1}^{t} \max\left\{0, \mathrm{val}_j(r^{j-1}) - \alpha \cdot \max_{z \in Z} E(u(z, \mathbf{A}_j(r^{j-1})))\right\}.$$

We now define what constitutes a good approximation of the first-best payoff.

Approximating the first-best. Fix two small numbers $\gamma, \delta > 0$. A strategy d of the DM achieves a (γ, δ)-approximation of the first-best payoff against a strategy a of the advisor if there exists a universal constant C (independent of $\mathbf{N}$) such that for every period t,

$$\Pr_{r^t \sim \mathbf{N}}\left[p^{DM,t} > \mathrm{FB}_t(r^{t-1}) - \max\{C, \gamma t\}\right] > 1 - \delta,$$

where $r^t \sim \mathbf{N}$ means that r_1 is drawn from $\mathbf{N}_1$, r_2 from $\mathbf{N}_2(r_1)$ and so on.

Thus, a strategy that achieves a (γ, δ)-approximation guarantees that with probability $1 - \delta$ (over the realizations of Nature), the difference between the average first-best payoff $\frac{\mathrm{FB}_t(r^{t-1})}{t}$ and the average payoff realized by the strategy is bounded above by γ after enough periods (specifically, after C/γ or more periods).[6]

The parameters γ and δ control the two types of challenges the DM faces when interacting with the advisor. The confidence parameter δ is needed since it may happen that even though the advisor knows Nature's process, Nature's realizations may be extremely unrepresentative, leading the DM to conclude that the advisor is uninformed. The accuracy parameter γ is needed because realized payoffs may differ in the short-run from expected payoffs.

In addition to achieving a good approximation of the first-best payoff against informed advisors, we would also like to make sure that the DM does not lose too much when trying to determine whether the advisor is informed.

Limited budget. A strategy d of the DM uses a realized budget of at most $m \in \mathbb{R}$ if for every strategy a of an advisor, every t, and every sequence of realizations r^t of Nature, it holds that

$$p^{DM,t} \geq -m.$$

Finally, it remains to define the notion of conservativeness.

ζ-Conservative advisor. Fix a small nonnegative number ζ. An advisor is ζ-*conservative* if any optimal action according to his information yields a weakly smaller or at most ζ-larger expected payoff under the advisor's process than under Nature's process. Formally, in every period t and sequence of realizations r^{t-1}

$$E(u(z_t, \mathbf{A}_t(r^{t-1}))) \leq E(u(z_t, \mathbf{N}_t(r^{t-1}))) + \zeta,$$

where $z_t \in \arg\max_{z \in Z} E(u(z, \mathbf{A}_t(r^{t-1})))$ denotes an optimal action according to the advisor's information.

Conservativeness arises naturally in the context of learning. If the advisor has a correct structural model of Nature (e.g., that Nature is an i.i.d. coin toss), yet needs to estimate the parameters of the model (e.g., the bias of the coin), then after

[6] Our notion of (γ, δ)-approximation is similar in spirit to the notion of Probably-Approximately-Correct (PAC) learning, in which a learner's goal is to predict with high probability (with respect to an unknown distribution) most of the decisions made by the agent who is the object of learning. See Kearns and Vazirani (1994) and Vidyasagar (1997) for the theory of PAC learning, and Kalai (2003), Salant (2007), Al-Najjar (2009), and Al-Najjar and Pai (2009) for applications of PAC learning to economics.

observing enough realizations of Nature's process, he will converge to Nature's process and hence be conservative. Example 1.2 demonstrates such a scenario.

Conservativeness also arises when the advisor tends to underestimate the likelihood of outcomes with positive payoffs and overestimate the likelihood of outcomes with negative payoffs. Consider, for example, a situation in which Nature's process is a sequence of coin tosses with outcomes in $R = \{-1, 1\}$ and that the possible bets $Z = \{-1, 1\}$ are on the direction of the coin. If the DM's payoff function has the form $u(z, r) = r \cdot z$ then being conservative amounts to correctly identifying the direction of the bias in every coin, yet weakly underestimating it. That is, either $\mathbf{N}_t(r^{t-1}) \geq \mathbf{A}_t(r^{t-1}) > 1/2$ or $\mathbf{N}_t(r^{t-1}) \leq \mathbf{A}_t(r^{t-1}) \leq 1/2$.

3.1 Possibility results

The next two theorems establish that there is a strategy D of the DM that achieves the following three payoff guarantees for every process of Nature and for every advisor. First, if the advisor is conservative and truthful (i.e., reports the maximal expected value and the payoff-maximizing action according to his information in every period), then D approximates the first-best payoff. Second, if the advisor is conservative and strategic, then the DM's payoff guarantee can only increase. Finally, the loss of the DM never exceeds a *fixed* threshold.

Recall that the per-period payoffs of the DM are in $[-a, b]$.

Theorem 3.1 *Let $\gamma > 2\zeta$ and let $k = O\left(\frac{-\log((\gamma-2\zeta)\delta)}{(\gamma-2\zeta)^2}\right)$.*[7] *There exists a strategy D of the DM such that for every process of Nature the following hold:*

- *The strategy D obtains a (γ, δ)-approximation of the first-best payoff against any truthful ζ-conservative advisor, where the constant C is $2k(a + b)$.*
- *The strategy D uses a realized budget of $m = k(a + \alpha b)$.*

Note that if the DM knew that the advisor is truthful and conservative, he could obtain the first-best payoff in expectation by simply following the advisor's recommendation in every stage. Theorem 3.1 establishes that it is possible to obtain with high probability a realized payoff that approximates the first-best payoff against a

[7]Formally, there exists a universal constant $B = B(a, b)$ such that for every γ, $\zeta < \gamma/2$, and δ, we have that $k \leq B \cdot \left(\frac{-\log((\gamma-2\zeta)\delta)}{(\gamma-2\zeta)^2}\right)$.

truthful conservative advisor even when the DM does not know ex-ante whether the advisor is truthful and conservative and by risking only a limited budget.

The strategy D that achieves the above guarantees is a simple modification of standard calibration strategies. It has two phases: In the first "test" phase, the DM buys information from the advisor whenever the advisor claims the value of the information is larger than some value β. In the second "calibration" phase, the DM buys information from the advisor whenever the advisor claims its value is larger than β and as long as the past recommendations of the advisor are within ϵ-distance from the payoff realizations of the DM. Setting $\epsilon = \frac{\gamma}{2}$ and $\beta = \frac{\epsilon}{1-\alpha}$ enables achieving the guarantees of Theorem 3.1. More formally,

DM's strategy. Suppose the interaction is currently in period t, the advisor's recommendations thus far have been $\{(v_j, z_j)\}_{j<t}$, the DM's actions have been $\{d_j\}_{j<t}$, and Nature's realizations have been $\{r_j\}_{j<t}$. The DM's strategy D in period t given the recommendation (v_t, z_t) is:

- If $\sum_{j<t} d_j < k$ then follow the advisor's prediction (i.e., set $d_t = 1$) if $v_t \geq \beta$.
- If $\sum_{j<t} d_j \geq k$ then follow the advisor's prediction if $v_t \geq \beta$ *and*

$$\frac{\sum_{j<t:d_j=1} u(r_j, z_j)}{\sum_{j<t} d_j} > \frac{\sum_{j<t:d_j=1} v_j}{\sum_{j<t} d_j} - \epsilon.$$

- Otherwise, set $d_t = 0$.

It is straightforward to verify that this strategy uses a realized budget of at most $m = k(a + \alpha b)$. Indeed, in the test phase, the DM purchases information at most k times, and whenever he does so, he pays the advisor at most αb and loses at most a. In the calibration phase, the payoff of the DM up to (and including) period t is

$$\begin{aligned}
\sum_{j<t:d_j=1} u(r_j, z_j) - \alpha \sum_{j<t:d_j=1} v_j - a - \alpha b \\
&\geq (1-\alpha) \sum_{j<t:d_j=1} v_j - \epsilon \sum_{j<t} d_j - a - \alpha b \\
&\geq ((1-\alpha)\beta - \epsilon) \sum_{j<t} d_j - a - \alpha b \\
&\geq -a - \alpha b,
\end{aligned}$$

where the first inequality is derived from the calibration condition (and the additional loss of $a+\alpha b$ is the worst possible outcome of round t), the second from the fact that the DM purchases information only when its value is larger than β, and the last from the choice of the parameters.

It is also straightforward to verify that if the DM did not "fail the advisor" – that is, if the calibration inequality of D is satisfied – then the DM's realized payoff approximates $\mathrm{FB}_t(r^{t-1})$. There are two cases to consider:

Case I: D is in the test phase. On the (at most) k periods in which the DM followed the advice, he lost at most $a+\alpha b$ per period. In addition, it holds that $\mathrm{FB}_t(r^{t-1}) \leq kb+\beta t$: the first term is the maximal value in the k periods in which the DM followed the advice and the second accounts for the $t-k$ periods in which the value of the advice was less than β. Thus, the payoff of the DM is at least

$$\begin{aligned} -k(a+\alpha b) &= (1-\alpha)(kb+\beta t) - k(a+b) - \beta t(1-\alpha) \\ &\geq \mathrm{FB}_t(r^{t-1}) - k(a+b) - \frac{\gamma t}{2} \\ &\geq \mathrm{FB}_t(r^{t-1}) - \max\{2k(a+b), \gamma t\}. \end{aligned}$$

Case II: D is in the calibration phase. The DM's payoff up to period t (as calculated above) is bounded below by

$$(1-\alpha) \sum_{j\leq t: d_j=1} v_j - \epsilon \sum_{j=1}^{t} d_j \geq (1-\alpha) \sum_{j\leq t: d_j=1} v_j - \epsilon t.$$

Since the DM buys information only when its value is $\geq \beta$, we have that

$$(1-\alpha) \sum_{j\leq t: d_j=1} v_j + (1-\alpha)\beta t \geq \mathrm{FB}_t(r^{t-1}).$$

Thus, the DM's payoff is at least $\mathrm{FB}_t(r^{t-1}) - (1-\alpha)\beta t - \epsilon t = \mathrm{FB}_t(r^{t-1}) - \gamma t$.

The challenging part of the proof, which appears in the appendix, is to show that the DM fails a truthful ζ-conservative advisor with probability at most δ, where the probability is over $r^t \sim \mathbf{N}$.

When being truthful, a ζ-conservative advisor can guarantee himself with probability $1-\delta$ a payoff of at least

$$V_t \stackrel{\text{def}}{=} \alpha \left(\sum_{j\leq t:\ \mathrm{val}_j(r^{j-1})>0} \mathrm{val}_j(r^{j-1}) - \beta t \right)$$

over t periods. This follows from the observation that if the advisor does not fail the test, his payoff is roughly an α-share of $\mathrm{val}_j(r^{j-1})$ in every period j in which $\mathrm{val}_j(r^{j-1}) \geq \beta$ and 0 otherwise. The probability that he fails the test is bounded above by δ.

Now suppose the advisor has a (mixed) strategy $\mathbf{S}$ that improves upon this payoff guarantee with high probability. The following result states that the strategy D still approximates the first-best payoff for the DM with high probability.

Theorem 3.2 *Suppose an advisor uses a strategy* $\mathbf{S}$ *for which*

$$\Pr_{\substack{r^t \sim \mathbf{N} \\ s \sim \mathbf{S}}} \left[p^{A,t}(s, D, r^t) > V_t + x\right] > 1 - \kappa$$

for some period t. *Then it holds that*

$$\Pr_{\substack{r^t \sim \mathbf{N} \\ s \sim \mathbf{S}}} \left[p^{DM,t}(s, D, r^t) > \mathrm{FB}_t(r^{t-1}) + \frac{(1-\alpha)x}{\alpha} - \max\{C, \gamma t\}\right] > 1 - \kappa,$$

where $C = 2k(a + b)$.

To summarize, the strategy D uses a finite and bounded budget. When interacting with a conservative and truthful advisor, the strategy achieves a payoff that approximates the first-best payoff with high probability. Any strategizing of a conservative advisor that improves his own payoff guarantee over truthfulness can only increase the payoff guarantee of the DM.

3.2 Impossibility result

We conclude this section by showing constructively that without conservativeness, it is impossible to achieve an approximation of the first-best payoff even when the advisor is informed. The reason is that approximating the first-best requires the DM to invest every once in a while, yet if the advisor is not conservative, then it is possible that whenever the DM invests, his payoff is small relative to the first-best.

Fix Nature's possible state realizations to be in $R = \{-1, 1\}$, the DM's possible bets to be in $Z = \{-1, 1\}$, and the DM's payoff function from investing to be $u(z, r) = r \cdot z$. That is, the DM obtains a payoff of 1 if he correctly predicts Nature's realization and -1 otherwise.

We allow the compensation scheme to the advisor to be any zero-liability scheme. We also allow the DM to have a larger set of actions: the DM can invest without consulting the advisor, and if he does consult with the advisor then he does not have to follow his advice. Thus, the DM's strategy in every period t is a probability distribution over pairs (d_t, a_t) where $d_t \in \{0,1\}$ specifies whether the DM consults with the advisor, and a_t specifies the action taken by the DM, which is either a bet in Z or "stay-out."

We now fix a strategy D' of the DM and show that D' does not achieve a good approximation of the first-best payoff against an advisor who is informed in the sense that the average per-period payoff from following his advice converges to the maximal attainable per-period payoff as the number of periods approaches infinity.

Consider a truthful advisor who believes Nature is a deterministic sequence $(A_1, A_2, \dots)$ of 1's and (-1)'s, and who is right from some period T that can be arbitrarily large. Nature is a deterministic sequence of 1's and (-1)'s as well. From period T onward, it is identical to the advisor's process. Up to period T, however, we construct Nature in a way that the DM obtains a payoff of at most 1/2 in every period while the first-best payoff is 1. Since T can be arbitrarily large, this implies that D' does not achieve the first-best payoff.

To construct Nature up to period T, let p_1^1, p_{-1}^1 and p_s^1 denote the probabilities (according to D') that the DM takes the action 1, -1 or stay-out in period 1. Nature's realization in period 1 is

$$r_1 = \begin{cases} A_1 & \text{if } p_s^1 > 1/2, \\ 1 & \text{if } p_s^1 \leq 1/2 \text{ and } p_1^1 - p_{-1}^1 \leq 1/2, \text{ and} \\ -1 & \text{otherwise.} \end{cases}$$

It is easy to verify that the expected payoff to the DM in period 1 is at most 1/2 (without even taking into account the payment to the advisor) while the first-best payoff is 1.

We now move to period 2. The probability with which the DM purchases information and with which he follows a particular action may depend on what he did in period 1 and on Nature's realization in case the DM invested. Since we know the distribution over the DM's actions, we can denote by p_1^2 the *marginal* probability that the DM takes the action 1 in period 2, and by p_{-1}^2 and p_s^2 the corresponding *marginal* probabilities of taking the action -1 and staying-out in period 2. We define r_2 in a similar fashion to r_1 using the marginal probabilities p_1^2, p_{-1}^2 and p_s^2.

We repeat this process until period T, getting a sequence $r_1, \dots, r_T$ on which the total expected difference of the DM's payoff from the first-best is at least $T/2$. Since

T is arbitrarily large, we obtain that:

Theorem 3.3 *Fix $\zeta < 1/2$. For every strategy D' of the DM and for every integer T, there exists a process of Nature and an advisor that is not ζ-conservative such that:*

- *The strategy D' does not obtain a (γ, δ)-approximation, for $\gamma < 1/2$, of the first-best payoff against the advisor until period T, even though*
- *the advisor is informed in the sense that the average per-period payoff from following his advice converges to the maximal attainable per-period payoff as the number of periods approaches infinity.*

Note that the same analysis extends to less restrictive compensation schemes in which the advisor partially compensates the DM for a loss (i.e., the advisor pays the DM some fraction of the realized loss that is bounded away from 1.)

4 Concluding comments

We conclude with two comments about possible generalizations of our results and the tightness of the approximation of the first-best payoff.

Arbitrary outside option. Throughout the paper we assumed that the DM's outside option is 0. A generalized framework would be one in which the outside option is allowed to be some $\theta \in \mathbb{R}$.

In the generalized framework, the definition of the first-best payoff is changed in a natural way. By staying out, the DM can obtain a payoff of θ in every period j. He should weigh that against the expected payoff of the best bet according to the advisor's information in period j, namely $\mathrm{val}_j(r^{j-1})$. Thus,

First-best payoff. The first-best payoff up to period t is

$$\mathrm{FB}_t(r^{t-1}) = \sum_{j=1}^{t} \max\{\theta, \mathrm{val}_j(r^{j-1}) - \alpha \cdot \max_{z \in Z} E(u(z, \mathbf{A}_j(r^{j-1})))\}.$$

Similarly, the definition of limited budget also changes. One may want to make sure that by interacting with the advisor, he does not lose more than some fixed amount with respect to what he could obtain by not interacting with the advisor, which is θt. Thus,

θ-Limited budget. A strategy d of the DM uses a θ-realized budget of at most $m \in \mathbb{R}$ if for every strategy a of an advisor, every t, and every sequence of realizations r^t of Nature, it holds that

$$p^{DM,t} \geq \theta t - m.$$

It is straightforward to extend Theorems 3.1, 3.2 and 3.3 to this generalized setting. For example, the modified version of Theorem 3.1 would state that there exists a strategy D_θ that achieves a (γ, δ)-approximation of the expected first-best payoff against any truthful ζ-conservative advisor such that the θ-realized budget of D_θ is $m = k(a + \alpha b)$. The strategy D_θ that achieves this guarantee is almost the same as the strategy D, except that in both the "test" and "calibration" phases the DM purchases information only if $v_t \geq \theta + \beta$ (as opposed to $v_t \geq \beta$ in D). The proof is essentially equivalent to that of Theorem 3.1, and is thus omitted.

Tightness of approximation. Three parameters govern the performance of the strategy D. The parameter γ specifies the desired distance from the first-best payoff. The parameter δ (or actually $1 - \delta$) specifies the probability with which the DM wishes to achieve this distance. Given the specification of these two parameters, a budget of $m = O\left(\frac{1}{\gamma^2} \log \frac{1}{\gamma\delta}\right)$ is sufficient to achieve a (γ, δ)-approximation (if $\zeta = 0$).

In practice, one may have a fixed budget m and may be interested in identifying which (γ, δ)-approximations are achievable with this budget. Our results indicate that using the strategy D, it is possible to achieve a $(\gamma, Be^{-\gamma^2 m}/\gamma)$-approximation for any $\gamma > 0$, where B is some universal constant. It remains an open question whether it is possible to achieve a tighter approximation. It also remains an open question whether tighter approximations can be achieved by relaxing the assumption on the messages sent by the advisor to the DM.

5 Appendix

5.1 Proof of Theorem 3.1

It remains to prove that the strategy D fails a truthful conservative advisor with probability at most δ. To prove this, we use a super-martingale inequality for random variables generated by a decision tree established by Chung and Lu (2006). We begin by describing Chung and Lu's (2006) setup and result, and then we show how to apply their result in our setting.

Let T be a tree[8] of depth n. For each node u in T, let $C(u)$ denote the finite set of the children[9] of u in T. The set $C(u)$ contains the possible outcomes given that the history of outcomes thus far has been the path from the root of T to u. For each edge from a node u to a node $v \in C(u)$, let p_{uv} be the probability of the outcome $v \in C(u)$ given that node u has been reached. Let f be a function from the nodes of T to $\mathbb{R}$.

Assume f satisfies the following properties:

- **Super-martingale:** For each node u it holds that $f(u) \leq \sum_{v \in C(u)} p_{uv} f(v)$.
- c**-Lipschitz:** For each u and $v \in C(u)$, it holds that $|f(u) - f(v)| \leq c$.
- σ**-bounded-variance:** For each node u it holds that

$$\sum_{v \in C(u)} p_{uv} f^2(v) - \left(\sum_{v \in C(u)} p_{uv} f(v) \right)^2 \leq \sigma^2.$$

Then,

Theorem 5.1 (Chung and Lu (2006), Theorem 8.8) *Fix a tree T of depth n and a function f satisfying the super-martingale, c-Lipschitz and σ-bounded-variance conditions. Let Y_0 be the root of T, and let Y_n be the random variable over the leaves of T generated by the tree process using the probabilities p_{uv}. Then*

$$\Pr\left[f(Y_n) \leq f(Y_0) - \lambda\right] \leq e^{-\frac{\lambda^2}{2\sigma^2 n + 2c\lambda/3}}.$$

Going back to our setup, we will now show that the strategy D fails a truthful ζ-conservative advisor with probability at most δ. It is sufficient to show that if the DM purchases information on recommendation (v_j, z_j) whenever $v_j \geq \beta$ then for any t and any ζ-conservative truthful advisor, it holds that

$$\Pr_{r^t \sim \mathbf{N}}\left[\frac{\sum_{j=1}^{s} d_j u(r_j, z_j)}{\sum_{j=1}^{s} d_j} \leq \frac{\sum_{j=1}^{s} d_j v_j}{\sum_{j=1}^{s} d_j} - \epsilon \text{ for some } s \in \{k, \ldots, t\}\right] \leq \delta.$$

[8] A tree is an undirected, rooted, acyclic graph. The depth of a node in a tree is the number of edges on the path connecting the node and the root, and the depth of a tree is the maximal depth of any node in the tree.

[9] A node v is a child of a node u if there is an edge connecting u and v, and if the path from the root to v passes through u.

The random variables generated by the stochastic process $\mathbf{N}$ naturally fit into the decision tree setup described above. A realization of Nature's process at any time period occurs with probability that depends on the history of realizations. The difficulty lies in the fact that the DM does not consider all periods meaningful, but only those that have a value at least β. The challenge is, then, to construct a tree (or rather, a forest[10]) and a function f that will allow us to use Theorem 5.1.

For the remainder of the proof, assume t is fixed. To simplify the construction, we add a 'dummy' period 0 realization r_0 to each possible realization $r^t = (r_1, \dots, r_t) \in R^t$, so the realizations are of the form $r^t = (r_0, r_1, \dots, r_t)$ where r_0 is identical across realizations. Since t is fixed, we also write r instead of r^t when it is clear from the context.

For each $s \in \{1, \dots, t\}$ and $r = (r_0, r_1, \dots, r_t)$, let $U_s(r)$ be the number of times in the first s periods the advisor's perceived value was at least β, assuming r is the sequence of realizations. Formally,

$$U_s(r) \stackrel{\text{def}}{=} \left|\{j : 1 \leq j \leq s \text{ and } v_j \geq \beta\}\right|,$$

where $v_j = v_j(r_0, \dots, r_{j-1})$ depends on past realizations in periods $0, \dots, j-1$.

Next, we define $t - k + 1$ sets of sequences of realizations, one for each $w \in \{k, \dots, t\}$, as follows:

$$R_w \stackrel{\text{def}}{=} \left\{r \in r_0 \times R^t : \exists s \leq t \text{ such that } U_s(r) = w\right\}.$$

Note that a specific sequence of realizations r may be in more than one of the sets R_w. A sequence r is absent from all R_w if the number of times in which $v_j \geq \beta$ is at most $k - 1$.

We now construct a forest F_w for a given w. The vertices of F_w correspond to the first w prefixes of sequences in R_w after which the value of information is at least β. Formally, the set of vertices is

$$\begin{aligned} U_w \stackrel{\text{def}}{=} \{(r_0, \dots, r_{j-1}) :\ & (r_0, \dots, r_{j-1}) \text{ is the prefix of some } r \in R_w, \\ & \text{and } v_j(r_0, \dots, r_{j-1}) \geq \beta, \\ & \text{and } \left|\{i < j-1 : v_i(r_0, \dots, r_{i-1}) \geq \beta\}\right| < w\}. \end{aligned}$$

The edges of F_w link every prefix to the minimal prefixes it nests. Formally, at each stage $i = 1, \dots, w$, we remove all vertices from U_w that have no prefix in U_w,

[10] A forest is a collection of disjoint trees.

and add them to level i of the forest F_w. We connect any vertex v removed in stage $i > 1$ to his parent, which is the unique vertex in level $i-1$ that is a prefix of v. Note that if $v_1 < \beta$ then F_w is a forest, and otherwise it is a tree.

We add transition probabilities between parents and their children in F_w as follows. For $u = (r_0, \ldots, r_\ell)$ and $v = (r_0, \ldots, r_\ell, \ldots r_j)$, where u is a parent of v, we assign to the edge connecting them the probability

$$p_{uv} = \Pr\left[\mathbf{N}_{\ell+1}(r_0, \ldots, r_\ell) = r_{\ell+1}\right] \cdot \ldots \cdot \Pr\left[\mathbf{N}_j(r_0, \ldots, r_{j-1}) = r_j\right].$$

We next modify F_w further by adding **final-realization nodes** as follows. For any vertex $u = (r_0, \ldots, r_{j-1}) \in U_w$ and any $r_j \in R$, let $p(r_j)$ be the total transition probability from u to its children whose j'th coordinate is r_j, i.e., $p(r_j) = \sum_{\substack{q \in C(u): \\ q_j = r_j}} p_{uq}$. If u is a leaf then $p(r_j) = 0$. If $(r_0, \ldots, r_{j-1}, r_j) \notin U_w$, we add the node $v = (r_0, \ldots, r_{j-1}, r_j)$ to the forest with transition probability

$$p_{uv} = \Pr\left[\mathbf{N}_j(r_0, \ldots, r_{j-1}) = r_j\right] - p(r_j).$$

Due to the addition of final-realization nodes, the transition probabilities from a non-leaf parent to its children now add up to one.

The resulting forest F_w has depth w (i.e., $w+1$ levels), but it may have leaves that are not at the bottom level. To remedy this we add **completion nodes** to F_w as follows. For any *leaf* u in F_w such that the depth of u is some $d < w$ we add $w - d$ nodes $u^{(1)}, \ldots, u^{(w-d)}$. The nodes are arranged in a line emerging from u: u is the parent of $u^{(1)}$, which is the parent of $u^{(2)}$, and so on until $u^{(w-d)}$. Transition probabilities between the new nodes are all 1.

The resulting forest F_w consists of trees of depth w, where the length of all paths from the root of a tree to a leaf is exactly w.

We now turn our attention to the second ingredient in the decision tree setup – the function f. We define f recursively with respect to the levels of the tree. First, set $f(\mathit{root}) = 0$ for the root of any tree in F_w. For any other node $v = (r_0, \ldots, r_j)$, let $u = (r_0, \ldots, r_\ell)$ be the parent of v in F_w (and so $\ell < j$). We set $f(v) = f(u)$ if v is a completion node, and otherwise,

$$f(v) = (u(z_{\ell+1}, r_{\ell+1}) - v_{\ell+1}) + f(u) + \zeta.$$

Intuitively, $f(v)$ sums up the differences between reported values and realized payoffs in those periods in which the reported value was at least β, leading up to but *not* including the realizations represented by v.

For any $r \in R_w$, let $L(r)$ be the leaf of the forest F_w that is reached when the sequence of realizations is r. The value $f(L(r))$ is then the difference between (i) the sum of realized values obtained by the DM given the realizations $r \in R_w$ plus w times ζ and (ii) the sum of expected values reported by the advisor. When the sequence r is chosen at random from $\mathbf{N}$, the random variable $f(L(r))$ captures the possible differences between realized values and expected values in the first w meaningful periods, conditional on $r \in R_w$.

The function f satisfies the following properties.

1. f **is a super-martingale.** If a vertex u is a completion node or a final realization node then its f value is identical to its only child's f value. For each vertex $u = (r_0, \ldots, r_j)$ that is neither a final realization node nor a completion node, it holds that

$$\begin{aligned}
&\left(\sum_{v \in C(u)} p_{uv} f(v)\right) - f(u) = \sum_{v \in C(u)} p_{uv} u(z_{j+1}, r_{j+1}) - v_{j+1} + \zeta \\
&= \left(\sum_{r_{j+1}} \left[u(z_{j+1}, r_{j+1}) \sum_{v \in C(u) \text{ s.t. } v_{j+1} = r_{j+1}} p_{uv}\right]\right) - v_{j+1} + \zeta \\
&= \left(\sum_{r_{j+1}} \Pr\left[\mathbf{N}_{j+1}(u) = r_{j+1}\right] \cdot u(z_{j+1}, r_{j+1})\right) - v_{j+1} + \zeta \\
&\geq 0,
\end{aligned}$$

where the final inequality follows from the fact that the advisor is truthful and ζ-conservative.

2. f **is** $(a+b)$**-Lipschitz.** This is an immediate implication of the fact that the range of the DM's utility function is $[-a, b]$.

3. f **has** $\left(\frac{a+b}{2}\right)$**-bounded-variance.** This is an immediate implication of the fact that the variance of each vertex is bounded above by the variance of a Bernoulli random variable that is 0 with probability 1/2 and $(a+b)$ with probability 1/2.

Theorem 5.1 implies that, for any λ,

$$\Pr_{r^t \sim \mathbf{N} \,:\, r^t \in R_w} \left[f(L(r^t)) \leq -\lambda\right] \leq e^{-\frac{\lambda^2}{w(a+b)^2/2 + 2\lambda(a+b)/3}}.$$

Fixing $\lambda = (\epsilon - \zeta)w$ yields

$$\Pr_{r^t \sim \mathbf{N} \,:\, r^t \in R_w} \left[f(L(r^t)) \leq -(\epsilon - \zeta)w \right] \leq e^{-\frac{((\epsilon-\zeta)w)^2}{w(a+b)^2/2 + 2(\epsilon-\zeta)w(a+b)/3}} \leq e^{-\frac{6(\epsilon-\zeta)^2 w}{5\max\{a+b,(a+b)^2\}}} \quad (1)$$

since $(\epsilon - \zeta) \leq 1$.

By the definition of f, the advisor fails in the calibration phase any time up to period t if for any $w \in \{k, \ldots, t\}$ it holds that $f(L(r)) \leq -(\epsilon - \zeta)w$ in the forest F_w for some realization r of length t with $U_t(r) \geq w$. We can use inequality (1) to calculate an upper bound on the probability that this happens by summing the right-hand expressions in this equation over all $w \in \{k, \ldots, t\}$. Denote by $B = \max\{a + b, (a + b)^2\}$. We obtain that the probability of failure is bounded above by

$$\sum_{w \in \{k,\ldots,t\}} e^{-\frac{6(\epsilon-\zeta)^2 w}{5B}} \leq \int_{k-1}^{\infty} e^{-\frac{6(\epsilon-\zeta)^2 w}{5B}} dw = \frac{5B}{6(\epsilon-\zeta)^2} \cdot e^{-\frac{6(\epsilon-\zeta)^2 w}{5B}} \Bigg|_{k-1}^{\infty} = \delta,$$

where the last equality holds for $k = \frac{5B}{6(\epsilon-\zeta)^2} \log \frac{5B}{6(\epsilon-\zeta)^2 \delta} + 1$.

5.2 Proof of Theorem 3.2

Fix a sequence of realizations $r^t = (r_1, \ldots, r_t)$, and let $\{(v_j^s, z_j^s)\}_{j \leq t}$ be the corresponding recommendations of the advisor playing $s \in \mathrm{supp}(\mathbf{S})$ when the DM plays D. Suppose that $p^{A,t}(s, D, r^t) > V_t + x$, and denote by

$$E_t(r^{t-1}) = \sum_{j \leq t:\ \mathrm{val}_j(r^{j-1}) > 0} \mathrm{val}_j(r^{j-1}).$$

There are two cases to consider:

Case I: D is in the test phase. It must be the case that

$$kb + \beta t > E_t(r^{t-1}) + \frac{x}{\alpha}. \quad (2)$$

To see why, observe that the DM bought information at most k times. Each time, he paid the advisor at most αb. Thus, $p^{A,t}(s, D, r^t) \leq \alpha k b$. By assumption it holds that $p^{A,t}(s, D, r^t) > V_t + x$. Combining the two inequalities yields (2).

In the (at most) k periods in which the DM followed the advice, he lost at most $a + \alpha b$ per period. Thus, the payoff of the DM is at least

$$\begin{aligned}
-k(a + \alpha b) &= (1 - \alpha)(kb + \beta t) - k(a + b) - \beta t(1 - \alpha) \\
&> (1 - \alpha)\left(E_t(r^{t-1}) + \frac{x}{\alpha}\right) - k(a + b) - \frac{\gamma t}{2} \\
&\geq \mathrm{FB}_t(r^{t-1}) + \frac{(1 - \alpha)x}{\alpha} - \max\{2k(a + b), \gamma t\}.
\end{aligned}$$

Case II: D is in the calibration phase. The payoff of the advisor up to period t is $p^{A,t}(s, D, r^t) = \alpha \cdot \sum_{j=1}^{t} d_j^s \cdot v_j^s$, where d_j^s denotes the action of the DM in period j given the realizations of Nature's process and the recommendations of the advisor up to period j. Since $p^{A,t}(s, D, r^t) > \alpha(E_t(r^{t-1}) - \beta t) + x$, it holds that

$$\sum_{j=1}^{t} d_j^s \cdot v_j^s > E_t(r^{t-1}) - \beta t + \frac{x}{\alpha}.$$

As shown in the main text, the payoff of the DM in the calibration phase is at least

$$(1-\alpha) \sum_{j=1}^{t} d_j^s \cdot v_j^s - \epsilon \sum_{j=1}^{t} d_j^s.$$

Substituting the left summation by $E_t(r^{t-1}) - \beta t + \frac{x}{\alpha}$ and the right summation by t, we obtain that the payoff of the DM is bounded below by

$$\begin{aligned}
(1-\alpha)\left(E_t(r^{t-1}) - \beta t + \frac{x}{\alpha}\right) - \epsilon t &= (1-\alpha)E_t(r^{t-1}) - t((1-\alpha)\beta + \epsilon) + \frac{x(1-\alpha)}{\alpha} \\
&= (1-\alpha)E_t(r^{t-1}) - \gamma t + \frac{x(1-\alpha)}{\alpha} \\
&\geq \mathrm{FB}_t(r^{t-1}) + \frac{x(1-\alpha)}{\alpha} - \gamma t.
\end{aligned}$$

Noting that r^t is a sequence of realizations for which $p^{A,t}(s, D, r^t) \geq \alpha(E_t(r^t) - \beta t) + x$ with probability $1 - \kappa$ over $r^t \sim \mathbf{N}$ and $s \sim \mathbf{S}$ completes the proof.

References

[1] N. Al-Najjar (2009). Decision Makers as Statisticians: Diversity, Ambiguity, and Learning. *Econometrica*, 77(5), 1339–1369.

[2] N. Al-Najjar and M. Pai (2009). Coarse Decision Making. *Mimeo.*

[3] N. Al-Najjar, A. Sandroni, R. Smorodinsky and J. Weinstein (2010). Testing Theories with Learnable and Predictive Representations. *Journal of Economic Theory*, 145(6), 2203–2217.

[4] N. Al-Najjar and J. Weinstein (2008). Comparative Testing of Experts. *Econometrica* 76(3), 541–559.

[5] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire (2002). The Nonstochastic Multiarmed Bandit Problem. *SIAM Journal on Computing*, 32(1), 48–77.

[6] N. Cesa-Bianchi and G. Lugosi (2006). Prediction, Learning and Games. Cambridge University Press.

[7] S. Chassang (2011). Calibrated Incentive Contracts. Mimeo.

[8] F. Chung and L. Lu. Concentration Inequalities and Martingale Inequalities: A survey (2006). *Internet Mathematics*, 3(1), 79–127.

[9] E. Dekel and Y. Feinberg (2006). Non-Bayesian Testing of a Stochastic Prediction. *Review of Economic Studies*, 73, 893–906.

[10] F. Echenique and E. Shmaya (2009). You Won't Harm Me if You Fool Me. Mimeo.

[11] D. P. Foster and R. Vohra (1999). Regret in the On-Line Decision Problems. *Games and Economic Behavior*, 29(1-2), 7–35.

[12] L. Fortnow and R. Vohra (2009). The Complexity of Forecast Testing. *Econometrica* 77(1), 93–105.

[13] H. Farnsworth (2011). Reputation and Portfolio Management Contracts. Manuscript.

[14] D. Fudenberg and D. K. Levine (1999). An Easier Way to Calibrate. *Games and Economic Behavior*, 29(1-2), 131–137.

[15] J. Hannan (1957). Approximation to Bayes Risk in Repeated Plays. In *Contributions to the Theory of Games, Volume 3* (M. Dresher, A.W. Tucker, and P. Wolfe, Eds.). Princeton University Press, Princeton, NJ.

[16] T.W. Hu and E. Shmaya (2010). Expressible Tests Need Not Be Manipulable. Manuscript.

[17] G. Kalai (2003), Learnability and Rationality of Choice, *Journal of Economic Theory*, 113(1), 104–117.

[18] M. J. Kearns and U. V. Vazirani (1994). An Introduction to Computational Learning Theory, MIT Press, Cambridge, MA.

[19] E. Lehrer (2001). Any Inspection Rule is Manipulable. *Econometrica*, 69(5), 1333–1347.

[20] W. Olszewski and M. Pęski (forthcoming). The Principal-Agent Approach to Testing Experts. *American Economic Journal: Microeconomics.*

[21] W. Olszewski and A. Sandroni (2008). Manipulability of Future-Independent Tests. *Econometrica*, 76(6), 1437–1466.

[22] W. Olszewski and A. Sandroni (2009). A Nonmanipulable Test. *Annals of Statistics*, 37(2), 1013–1039.

[23] Y. Salant (2007). On the Learnability of Majority Rule. *Journal of Economic Theory*, 135(1), 196–213.

[24] A. Sandroni (2003). The Reproducible Properties of Correct Forecasts. *International Journal of Game Theory*, 32(1), 151–159.

[25] A. Sandroni, R. Smorodinsky and R. Vohra (2003). Calibration with Many Checking Rules. *Mathematics of Operations Research*, 28(1), 141–153.

[26] E. Shmaya (2008). Many Inspections are Manipulable. Theoretical Economics, 3(3), 367–382.

[27] M. Vidyasagar (1997). A Theory of Learning and Generalization, Springer, London.

[28] V. Vovk and G. Shafer (2005). Good Randomized Sequential Probability Forecasting is Always Possible. *Journal of the Royal Statistical Society: Series B*, 67(5), 747–763.