Keeping Experts Honest*

Ricardo Fonseca†

October 14, 2023

Abstract

Decision-makers often need advice from specialists, who can offer a more precise assessment of the state of the world, but also be biased towards one action. One such situation is the decision to fund scientific research made by NIH/NSF officers with the help of a peer review system. The present paper studies how a principal can fight these types of biases by committing to future decisions that affect the payoff of the expert, in an environment in which transfers are not allowed. In particular, it establishes that a particularly simple type of mechanism that works by randomizing "Grim-trigger"-type strategies, in which a principal never again listens to an expert after a bad recommendation, is optimal.

Keywords: Mechanism Design; repeated games; information transmission

JEL Codes: C73, D82

*I am grateful to Roberto Serrano's guidance and comments from Jack Fanning, Teddy Mekonnen, Bobby Pakzad-Hurson, Kareen Rozen, and Theory Lunch at Brown University participants.

†Brown University, Economics Department. 8 Fones Alley, Providence, R.I., 02912. E-mail: ricardo_fonseca@brown.edu.

1 Introduction

Decision-makers often rely on the opinions of experts. Among many, one such situation is the process of funding scientific research: an agency (e.g., the NSF or the NIH) uses a peer review system to evaluate projects that might receive grants[1]. It is the case that the interests of the experts are often not wholly aligned with those of the decision-maker. A scientist might be biased against the adoption of a particular novel procedure and recommend against the funding of a project using it. .

Suppose that the decision-maker has to repeatedly make decisions depending on a series of uncertain states of the world and cannot make conditional transfers to the experts. If he can commit to a decision function based on advisors' recommendations, varying with time only through performance, but using a pre-committed algorithm, how should he optimally do it?

This paper presents a discrete-time infinite horizon principal-agent model with uncertainty regarding a sequence of states of the world. The long-lived agents, referred to as the "experts" here, have signals with different precision on the state of the world at each period. They prefer funding good projects but might be biased toward funding even bad ones, so their preferences are not entirely aligned with the principal's. One way to motivate this assumption is to think of the expert as wanting to see a particular type of project, no matter its quality, being carried through. In the first example above, one can think of a scientist who not only wants to fund promising research but also wants to see projects using a certain methodology dominate a field.

[1]There is literature presenting evidence that the funding process from the NIH (National Institute of Health), based on peer review, has been mainly funding incremental research, not breakthroughs (e.g. Azoulay et al. (2011), Stephan (2012)). To the extent that the bias from peer review evaluators (see, for example, Li (2017) for evidence of bias in NIH peer review) drives this, the present study discusses how to deal with it optimally. More recently, Bhattacharya and Packalen (2020) argued that there are general problems with scientific production today, in particular, concerning the pace of breakthroughs' discovery, and proposes changes to the argued exaggerated emphasis on paper citations as a way to fight it. For a recent overview of the argument for scientific grants and suggestions on how to design more efficient systems of scientific research funding, see Azoulay and Li (2020)

Suppose that the principal cannot make transfers to the experts, but can change the decision function as a tool to reward them for good advice in the past. For example, he can announce that the weight of an expert on the decision-making function for future periods will increase if she develops a good recommendation track record.

Our main result shows that a variation of a Grim-trigger strategy, denoted by AUNT (Almost Unforgiving Information Transmission mechanism), in which the principal commits to, with some probability ϵ, never funding any project (an action that is least preferred by any type of agent) forever after a bad recommendation by the expert is optimal. The value of ϵ is chosen to make the agent just indifferent between telling the true signal realization or not when it indicates that the project is of low quality.

The set of potential mechanisms that the principal can make is potentially large. One class that is of particular interest is comprised of those in which the principal can do the least preferred action for the adviser by some K periods and then go back to listening to the expert. The payoffs that a planner can achieve by using an AUNT mechanism are the highest among all possible ones.

The present paper is structured as follows: in Section 2, we discuss a literature review of related work; in Section 3, the general model is introduced; in Section 4 and Section 5, we analyze the first-best and introduce our optimal mechanism. Lastly, Section 6 concludes.

2 Literature Review

There are three main strands of the economic theory literature related to this work: dynamic mechanism design with no transfers, the literature on how to evaluate expert performance, and also a collection of papers focused on the problem of how to aggregate expert opinions. Besides that, there is empirical literature in the subfield of economics of science on the problems with science funding in recent periods.

The literature on dynamic mechanism design with no transfers is relatively recent and most closely connected to this work. Li et al. (2017) studies the problem of how to create incentives for a single agent to report the known best project for the principle, leading to promises of more authority for the agent after accurate recommendations, shedding light on why some organizations may end up giving managers substantial power in response to good performance in the first periods. Guo and Hörner (2018) study the problem of efficiently allocating perishable goods across time, when the agent has a private valuation in a binary set, following a Markov state transition. They show that a system with characteristics similar to a quota one (Jackson and Sonnenschein (2007)), but in which the quota varies with the agent's action, achieves the efficient allocation. Meng (2018) drops the assumption of a persistent state, considering i.i.d. ones, but generalizes the state space, actions, and preferences, achieving a "Folk Theorem" result, together with identifying the rate of convergence. Lipnowski and Ramos (2019) studies a problem of optimal delegation in a setting without principal commitment power, leading to an optimal system in which the agent loses autonomy over time. All the papers above consider models in which the principal does not learn about the accuracy of the agent reports, and there is only one agent, who can perfectly see the state of the world, unlike the present project.

De Clippel et al. (2019) and Margaria and Smolin (2018), also part of this literature, study models of asymmetric information with many agents, but again the agents can see the state of the world perfectly. The first paper also assumes no commitment power, and the second focuses on a setting where the principal cannot learn the accuracy of previous reports from previous periods.

There is also literature on how to select potentially biased experts for advice: Gerardi and Yariv (2008), Che and Kartik (2009), Bhattacharya and Mukherjee (2013). The focus of these papers is more on selecting experts from a rich pool with the costly acquisition of information on project quality, and their models are static. Deb et al. (2018) have a dynamic model with commitment but focus on the best timing to evaluate a forecaster to learn about her precision. In contrast, our model focuses on the dynamic aspects of the process of using expert opinions, with the aim of fighting biases, not learning about precision, which is assumed to be known from the start, or inducing effort.

There is also a small literature on aggregating expert opinions (Glazer and Rubinstein, 1998; Gerardi et al., 2009; Pakzad-Hurson, 2021) focusing on static models and biases in terms of the reputations of experts. In comparison, the experts in our model have an interest in a state-independent specific action.

Literature in finance studies the problem of motivating innovation (e.g. Manso (2011), Balsmeier et al. (2017)). The main takeaway is that early flexibility is important for risk-taking. Unlike this literature, we are not considering the incentives of researchers to explore more risky methods, but instead on how to fight expert evaluator biases and induce them to report truthfully.

More related to the research funding setting in particular, Azoulay et al. (2011) shows that a program of research funding focused on researchers instead of projects, with more leniency on early failure, has better results in terms of producing high-impact, breakthrough research. Although this might be a relevant tool, with interesting implications, it can hardly be applied systematically by an organization like the NIH or NSF, which has to deal with research done by scientists with varied levels of seniority.

3 Model

3.1 Structure

Consider a model with an infinite discrete-time horizon, one long-lived principal ("he"), and one long-lived advisor. We also refer to the advisor as "agent" or "expert" henceforth. At each time t, a new state of the world $q_t \in \{0, 1\}$ materializes. The "q_t" here stands for project quality at time t, which can be high or low, 1 or 0, respectively. The probability that $q_t = 1$ equals q, and the realizations are i.i.d over time. The advisor can see a conditionally independent signal $\eta_t \in \{0, 1\}$ every period of time, with $Pr(\eta_t = q_t | q_t) = p$. Without loss of generality, we can assume that $p > 1/2$, as otherwise, one can take a realization as evidence that the state of the world is the opposite of the result. We denote by p the advisor's precision, exogenously determined and commonly known by all, and therefore independent of the expert's effort.

After seeing the signal realization, the advisor writes a recommendation $r_t \in \{0, 1\}$, where $r_t = 1$ is, without loss of generality, a recommendation to fund the project presented at time t. After seeing the recommendation, the principal must take a decision on whether to fund the project with some probability $\rho^r(h^t)$ for any public history $h^t = ((r_1, q_1, x_1), (r_2, q_2, x_2)..., (r_{t-1}, q_{t-1}, x_{t-1}))$ and a recommendation report r, where x_t is a public randomization device, that takes values on $U[0, 1]$. This history is public, as all of the realizations are known by the players at time t. Denote the set of all histories as $\mathcal{H}$, and the set of all histories consistent with a $\hat{h}^t$ for some period t by $\mathcal{H}(\hat{h}^t)$.

In every period $t + 1$, whether the project was funded or not, q_t is revealed, and the payoffs are attained. We will see later that the space of preferences is such that the optimal mechanism will require only that the quality of funded projects should be observed.

At time $t = 0$, the principal can commit to a decision mechanism d. Before the advisor decides at each period, a draw from the public randomization device x_t is revealed to all. The mechanism specifies the $\rho^r(h^t)$ probability of funding the project for every public history h^t and every report r. Given $\rho^{r'}(\hat{h}^t)$, we have that the project is funded, or $d_t = 1$, after history $\hat{h}^t$ and a report r' with probability $\rho^{r'}(\hat{h}^t)$ and not funded, or $d_t = 0$, with the complementary probability. The strategies of the principal are then limited to choosing a d decision mechanism at time 0 and committing to it thereafter.

The agents have reporting actions $r_t(d, h^t, \eta_t) \in \{0, 1\}$, so that they choose a recommendation conditional on every public history h^t, decision mechanism d and signal realization η_t, on whether to fund the project presented at time t ($r_t = 1$) or not ($r_t = 0$). Reporting strategies are then randomized on these actions. Formally, $\sigma_t(\hat{d}, \hat{h}^{t'}, \hat{\eta}_{t'}) \in [0, 1]$ denotes the probability of recommending that the project be funded after history $\hat{h}^{t'}$ on time t' and observing a signal realization $\hat{\eta}_{t'}$, with mechanism $\hat{d}$ chosen at time 0. Note that as the signals are i.i.d, there is no reason to condition the strategies on the private history of signal realizations.

3.2 Preferences

The principal gets a utility of 1 if he funds a project with $q_t = 1$, -1 if he funds a project with $q_t = 0$, and 0 if he does not fund the project, regardless of quality. The period utility attained after a funding decision and project quality at time t is represented by $u_P(d_t, q_t)$.

The principal discounts payoffs at rate $\delta \in (0, 1)$. Therefore he commits to a plan d that solves:

$$U^P(d) = (1 - \delta)E\left[\sum_t \delta^t u_P(d_t, q_t)\right]$$

This expectation is taken with respect to the possible realization of the stream of project qualities and signals, together with the mixed strategies and realizations of the public randomization device.

The expert has period utility represented by the following formula:

$$u(d_t, r_t, q_t) = \lambda u_P(d_t, r_t, q_t) + (1 - \lambda)\mathbb{1}_{\{d_t=1\}}$$

The expert has a payoff of a weighted average of the principal's utility and an extra amount that comes from getting a project funded, regardless of quality. This last factor represents a potential bias toward recommending that projects be funded. If $\lambda = 1$, $u_i(.) = u_P(.)$, so their preferences are completely aligned. If $\lambda = 0$, the agent cares only about funding projects, regardless of quality. The value of the parameter λ is common knowledge for all players.

The expert tries to maximize the expected total discounted utility from period t forward, all with a common discount factor δ, which is the discount factor of the principal as well:

$$U_t(d_t, \eta_t, h^t) = (1-\delta)E\left[\sum_{t' \geq t} \delta^{t'} u(d_{t'}, r_{t'}, q_{t'})\right]$$

This equation indicates the agent's payoff at time t after observing a signal η_t, knowing that the principal chose the mechanism d (and therefore makes decisions d_t, potentially after randomizing), and the public history $h^t = ((r_1, q_1, x_1), (r_2, q_2, x_2)..., (r_{t-1}, q_{t-1}, x_{t-1}))$. The expectation is taken with respect to future realizations of project qualities, random decision-making, future signal realizations, and the future realizations of the public randomization device.

The payoffs are multiplied by $(1-\delta)$, as the principal's payoff. Unlike the principal, they do not have commitment power and must maximize at time t their utility from that period forward. We will assume that if indifferent between two reports, the agent breaks ties by reporting her signal realization[2].

Note that there is only one direction for the bias: towards funding them no matter what. The experts are not allowed to have a bias towards not funding projects, for example. This is a minor point with one agent, as if the bias was towards not funding any project, we would change the optimal mechanisms presented in a straightforward manner. It does have significant implications for the analysis with many agents, though.

Our focus is on the perfect Bayesian equilibria (PBE) of this game: the principal picks a

[2]This assumption is not strictly necessary for the results and is made uniquely for simplicity.

mechanism d that maximizes $U^P(d)$, her expected payoff at time 0, while each agent chooses a reporting strategy $\sigma_t(d, \hat{h}^t, \hat{\eta}_t) \in [0,1]$ maximizing its expected utility $u(d, r_t, q_t)$ from each period t forward, after history $\hat{h}^t$ and signal η_t.

4 The Principal's First Best

For this section, our objective is to identify an optimal mechanism for the principal (there are potentially multiple). The discussion in this section is applicable for any λ. So the points raised should apply even if the principal does not know how biased the agent is.

Let's start by defining the principal's first-best:

Definition 1 *A mechanism d reaches the **principal's first-best** if the principal gets the same expected utility from it as if he could observe the signal realizations directly.*

Note that this is not a welfare notion considering both players' utility. The principal's first-best is a benchmark for what payoff the principal can hope to achieve.

One first thing to note is that if $q \leq 1-p$ or $q \geq p$, the principal would just ignore the signal realization. In the former case, it is better for him not to fund any project, even if $\eta_t = 1$. In the latter, he should fund the project anyway, even if the signal realization is suggesting that the project is bad. Our focus will therefore be on cases in which $1-p < q < p$, so that the expert has information that is useful for the principal's decision.

Given our condition on q, and the precision of the signal, the first-best payoff for the principal is equal to:

$$[Pr(q_t = 1|\eta_t = 1) - Pr(q_t = 0|\eta_t = 1)]Pr(\eta_t = 1)$$

as he will always prefer to follow the signal and get a payoff of 1 if the project is good and funded, -1 if it is bad and funded, and 0 if it is not funded at all.

This leads to the following payoff at time t:

$$\left(\frac{Pr(\eta_t = 1|q_t = 1)Pr(q_t = 1)}{Pr(\eta_t = 1)} - \frac{Pr(\eta_t = 1|q_t = 0)Pr(q_t = 0)}{Pr(\eta_t = 1)} \right) Pr(\eta_t = 1) = p + q - 1 \geq 0$$

Where the inequality holds by our assumptions on q and p. The first mechanism that we will introduce essentially uses a Grim-Trigger strategy by the principal, forever punishing the agent after a bad report through the commitment to never fund a project again.

Definition 2 *The **Unforgiving Information Transmission Mechanism (UNIT)**, denoted by d^{UM}, is characterized by $d_t(h^t, r_t) = r_t$ if $\forall t' \leq t$, it is not the case that $r_t = 1$ and $q_t = 0$, and $d_t(h^t, r_t) = 0$ otherwise.*

Note that when the agent has a perfectly informative signal, so that $p = 1$, and the agent is patient enough [3], UNIT can make the principal achieve the first-best. Subsection 6, in the Appendix, shows that this is robust to almost-perfect precision, in the sense that if p is very close to 1, the principal can get to payoffs arbitrarily close to first-best by using the UNIT mechanism. The situation is radically different when $p < 1$, as will be discussed shortly.

4.1 The Impossibility of Achieving The First-Best

The positive result in the case of perfect precision fundamentally relies on the fact that the agent knows the state of the world for sure, and has no hope of holding whenever $p < 1$. Let's define the following mechanism, characterized by always following whatever the expert recommends:

Definition 3 *The **Rubber-stamping** mechanism is characterized by $d_t(h^t, r_t) = r_t$ at any t, for any history h^t and any recommendation profile r_t.*

[3] Formally, for $p = 1$ and $\delta \geq 1/1 + q$, the UNIT Mechanism achieves the first-best for any λ. The important case to check is when $\eta_t = 0$. For there not to be a deviation, we need to have $(1-\delta)(1-2\lambda) \leq \delta q$, with the left-hand side giving the agent's expected utility from $r_t = 1$.

Let's denote the rubber-stamping mechanism as d^{RS}. A natural question is what is the principal's payoff from choosing it as the decision-making mechanism, for a given p, λ, and q?

At every t, if $\eta_t = 1$, $r_t = 1$ is optimal for the agent, as reporting $r_t = 0$ leads to a lower payoff today and does not change future payoffs. If $\eta_t = 0$, the agent gets $\lambda(Pr(q_t = 1|\eta_t = 0) - Prob(q_t = 0|\eta_t = 0)) + 1 - \lambda = \frac{(1-p)q+(1-2\lambda)p(1-q)}{(1-p)q+(1-q)p}$ by picking $r_t = 1$ and 0 by picking $r_t = 0$. If $\lambda < \frac{(1-p)q+(1-q)p}{2(1-q)p} \equiv \lambda(p,q)$, it is always better for her to lie, and rubber-stamping leads to the agent picking $r_t = 1$ every time, which leads to $U^P(d^{RS}) = 2q - 1 < p + q - 1 = U^{FB}$, as $p > q$ by assumption. We have, then, that $U^P(d^{RS}) = U^{FB}$ if and only if $\lambda \geq \lambda(p,q)$.

The following result tells us that if rubber-stamping does not achieve the first-best, no mechanism can, when $p < 1$.

Theorem 1 *Consider the case in which $p < 1$. If $\lambda < \lambda(p,q)$, no mechanism can achieve the first-best.*

Proof. As we know from the discussion above, when $\lambda < \lambda(p,q)$, rubber-stamping cannot achieve the first-best. Suppose that another mechanism M can. If it just does the opposite of what the agent recommends, it is equivalent to rubber-stamping, so it also does not achieve the first-best. Suppose now that it follows one of the two possible reports with positive probability. As it is not rubber-stamping or its equivalent, it must listen to one of the reports with a probability less than 1 or higher than 0. But then it cannot fully use the value from the signals and therefore cannot reach the first-best. ■

Note that the above result is easily generalizable for the case with I experts with $p_i < 1$ for all experts, the exact complement of the multi-agent extension to the previous theorem. This result tells us that there is no hope of achieving the first-best whenever the expert has imperfect precision and rubber-stamping cannot do it.

In the next section, we will see that even though the first-best cannot be achieved when $p < 1$, a mechanism using a similar punishment as UNIT with some probability is optimal.

5 An Optimal Mechanism

Let's keep for now the assumption that $I = 1$, so that there is only one advisor, once again. We will consider the following mechanism: if the expert recommends funding a project that is shown to be bad, with probability ϵ the principal will never listen to her again. Note that this is a restrictive form of punishment for the expert. One could think of mechanisms where the principal does not listen to the advisor for K periods, or randomizes every time, among others. Our main result in this section is that focusing on mechanisms of this type is without loss of generality.

Let's define this class of mechanisms and find the best ϵ for the principal as a function of λ, denoting it by $\epsilon(\lambda)$.

Definition 4 *Define the* ***Almost-Unforgiving Information Transmission mechanism (AUNT)*** *as the one that rubber-stamps agent recommendations but, with probability ϵ, never funds any project again after a wrong recommendation to fund a project.*

Note that $\epsilon = 1$ turns it into the UNIT mechanism, and $\epsilon = 0$ makes it equivalent to the rubber-stamping mechanism. This random probability of punishing the agent or not is the reason why we need the public randomization device x_t.

Note that AUNT Mechanisms have many desirable features. For example, they are simple and Markovian, having the same rules at each point in time. Note also that they do not depend on the principal's ability to observe the quality of projects that were not funded. This is reassuring, as we can bypass the assumption that even unfunded projects can have their quality revealed. A straightforward modification of it, reversing incentives, can deal with the case in which the agent has a bias towards not funding projects: after a misreport of telling that a good project is bad, fund projects forever with some probability. This way, all the incentives are reversed in the same faction as before.

Let's check what is the optimal ϵ for a certain λ, denoted by $\epsilon(\lambda)$ here. We will first assume that the mechanism must "keep the agent honest", which means that the agent never misreports her signal realizations. We will then argue that AUNT mechanisms with $\epsilon = \epsilon(\lambda)$ are optimal.

Denote by $\sigma_R \geq 0$ the "reward" continuation value, after a correct recommendation, after any history in which the principal did not already commit to never fund any project. To keep the agent honest when $\eta_t = 0$, we need:

$$\sigma_R \geq \frac{(1-\delta)[q + (1-2\lambda)p - 2pq(1-\lambda)]}{\delta\epsilon(1-q)p}$$

This inequality comes from comparing the payoffs from each recommendation and requiring that the one from picking $r_t = 0$ be greater or equal to the one from picking $r_t = 1$.

To keep the agent honest when $\eta_t = 1$, we need:

$$\sigma_R \leq \frac{(1-\delta)(pq + (1-2\lambda)(1-p)(1-q))}{\delta\epsilon(1-p)(1-q)}$$

The promise-keeping constraint leads to, for $\kappa = pq + (1-q)(1-p)$:

$$\sigma_R = \frac{(1-\delta)(\kappa - 2\lambda(1-q)(1-p)}{1-\delta(1-\epsilon(1-q)(1-p))}$$

One can see that the truth-telling constraint for $\eta_t = 1$ is satisfied for any δ, then, and we can therefore ignore it.

Rearranging the truth-telling constraint for $\eta_t = 0$, we get the following condition for ϵ:

$$\epsilon \geq \frac{(1-\delta)[1-\kappa-2\lambda p(1-q)]}{\delta(1-q)(2p-1)q}$$

To have $\epsilon \leq 1$ for any λ, we need $\delta \geq [1-\kappa]/[p(1-q)+q(1-\kappa)]$. Informally, if δ is high enough, the $UNIT$ mechanism induces a punishment that is strong enough to guarantee truth-telling when $\eta_t = 0$. With $\epsilon = 0$, AUNT is just rubber-stamping, that gives the highest feasible payoffs for the agent. As the principal wants to give the lowest possible punishment in order to keep the agent honest, he makes her indifferent between truth-telling or not after a signal realization indicating that the project is bad.

We have that the optimal ϵ is, then:

$$\epsilon(\lambda) = \begin{cases} 0 \text{ if } \lambda \geq \lambda(p,q) \\ \frac{(1-\delta)[1-\kappa-2\lambda p(1-q)]}{\delta(1-q)(2p-1)q} \text{ if } \lambda < \lambda(p,q) \end{cases}$$

The payoff for the principal is equal to:

$$U^P(d) = \frac{(1-\delta)(p+q-1)}{(1-\delta)+\delta\epsilon(1-q)(1-p)}$$

Using the optimal ϵ from above, we get that

$$U^P(d) = \begin{cases} p+q-1 \text{ if } \lambda \geq \lambda(p,q) \\ \frac{(2p-1)q(p+q-1)}{p(pq+(1-2\lambda)(1-p)(1-q))} \text{ if } \lambda < \lambda(p,q) \end{cases}$$

The $\epsilon(\lambda)$ above will characterize the best mechanism for the principal in the class of AUNT mechanisms. The next result will tell us that restricting ourselves to mechanisms in this subclass is without loss of generality, in the sense that the AUNT mechanism with $\epsilon = \epsilon(\lambda)$ is optimal.

Theorem 2 *The AUNT mechanism with $\epsilon = \epsilon(\lambda)$ is optimal for $\delta \geq 1/(q+1)$.*

As a preliminary step, note first that if $\lambda \geq \lambda(p,q)$ the preferences of principal and agent are aligned enough to make Rubber-Stamping, a special case of AUNT (with $\epsilon = 0$) hold. If $\lambda < \lambda(p,q)$ we know that the agent has a preference for $d_t = 0$ for every signal realization. We can therefore focus on this case for the remainder of our argument.

We will now prove that a version of the Revelation Principle holds in our setting. In particular, we need to formally establish that it is in the best interest of the principal to make the agent report her observed signal realizations. This can be done by using a very simple argument, though, laid out below

Lemma 1 *There is an optimal mechanism d^* that induces the agent to report that $r_t = \eta_t$ in order to maximize $U_t^i(d^*, \eta_t, h^t)$ after any history h^t and realization η_t.*

Proof. Take an optimal $\tilde{d}$ and a history $\tilde{h}^t$ after which the agent finds it in her best interest to misreport her signal. The principal can use d^* instead, which acts in the same way as $\tilde{d}$ right after $\tilde{h}^t$ for any report and promises to treat the agent in the same way as $\tilde{d}$ from period $t+1$ forward.

To see this more clearly, suppose that, for $\tilde{d}$, after $\tilde{h}^t$ the agent's payoff from reporting $r_t = 1$ is greater than the one from reporting $r_t = 0$ for any signal realization. Then the principal can promise, for a new mechanism d^*, to implement the recommendation with the same probability as in $\tilde{d}$ for $r_t = 1$ for any report $\rho^1(\tilde{h}^t) = \rho^0(\tilde{h}^t)$ and to implement future recommendations with the same probability as in $\tilde{d}$. Therefore $\sigma_R^0(\tilde{h}^t) = \sigma_R^1(\tilde{h}^t)$ and $\sigma_P^0(\tilde{h}^t) = \sigma_P^1(\tilde{h}^t)$ and truth-telling is optimal for the agent (as any report leads to the same payoff for her). As the agent and principal act in the same way as in $\tilde{d}$, d^* is also optimal for the principal. Analogous arguments hold when the equilibrium reports by the agent are the opposite of the signal realization or $r_t = 0$ for any realization. As the principal can do this for any period and history in which misreporting is optimal, there is an optimal mechanism in which the agent always reports $r_t = \eta_t$ in equilibrium. ■

By the lemma above, we can focus, without loss of generality, on mechanisms that keep the agent honest. We call any such mechanism truthful.

The next lemma will tell us that the agent should not be punished for a recommendation not to fund a good project. The crucial point is to note that given the optimal truth-telling behavior by the agent, implementing her recommendations with a higher probability in the future is a tool that the principal has to increase his future payoffs. By following the recommendation to implement a recommendation to fund with a higher probability after some history, he is increasing the payoffs of the agent. By following a recommendation not to fund after some history, he is decreasing her payoff, on the other hand, as $\lambda < \lambda(p, q)$.

Lemma 2 *There is a truthful optimal mechanism $\tilde{d}$ inducing $\tilde{\sigma}^0_R(h^t) = \tilde{\sigma}^0_P(h^t)$ after any history h^t.*

Proof. Suppose that there is a truthful optimal mechanism $\hat{d}$ such that $\hat{\sigma}^0_R(\hat{h}^t) > \hat{\sigma}^0_P(\hat{h}^t)$ after some history $\hat{h}^t$. Given the fact that $\hat{d}$ is a truthful mechanism and that the agent breaks ties by making recommendations in line with her signal realization, we know that the truth-telling constraints hold:

$$p(1-q)\hat{\sigma}^0_R(\hat{h}^t) + q(1-p)\hat{\sigma}^0_P(\hat{h}^t) \geq q(1-p)\hat{\sigma}^1_R(\hat{h}^t) + p(1-q)\hat{\sigma}^1_P(\hat{h}^t)$$

after seeing $\eta_t = 0$ and

$$pq\hat{\sigma}^1_R(\hat{h}^t) + (1-p)(1-q)\hat{\sigma}^1_P(\hat{h}^t) \geq pq\hat{\sigma}^0_P(\hat{h}^t) + (1-p)(1-q)\hat{\sigma}^0_R(\hat{h}^t)$$

after observing $\eta_t = 1$.

It must be the case that after two future histories $\bar{h}^{t'}, \tilde{h}^{t''} \in \mathcal{H}(\hat{h}^t)$, differing only in the state of the world q_t at $t < t', t''$ but following $\hat{h}^t$ before t, either the principal is implementing $r_{t'} = 1$ with higher probability after a recommendation $r_t = 0$ that matches the state than after a wrong one, or following a recommendation $r_{t'} = 0$ with lower probability, as otherwise we would have $\hat{\sigma}^0_R(\hat{h}^t) = \hat{\sigma}^0_P(\hat{h}^t)$[4].

Consider the former case first. Suppose that the principal were to instead use $\tilde{d}$, that after $\hat{h}^t$ and a misreport $r_t = 0$, follows the advice $r_{t'} = 1$ after $\tilde{h}^{t'}$ with some higher probability, and follows $\hat{d}$ in any other period and after any other history in $\mathcal{H}(\hat{h}^t)$, and therefore induces a $\tilde{\sigma}^0_P(\hat{h}^t)$ in the open interval $(\hat{\sigma}^0_P(\hat{h}^t), \hat{\sigma}^0_R(\hat{h}^t))$. If $\tilde{d}$ keeps the truth-telling constraints holding, it is a better mechanism for the principal, as it follows the recommendation matching the signal realization (by truth-telling) more often.

The only possibility of not being able to find such $\tilde{\sigma}^0_P(\hat{h}^t)$ as described is if the truth-telling constraint after observing $\eta_t = 1$ binds for $\hat{d}$. If so, then:

$$pq\hat{\sigma}^1_R(\hat{h}^t) + (1-p)(1-q)\hat{\sigma}^1_P(\hat{h}^t) = pq\hat{\sigma}^0_P(\hat{h}^t) + (1-p)(1-q)\hat{\sigma}^0_R(\hat{h}^t)$$

after observing $\eta_t = 1$.

By feasibility, we have that, for $\hat{\sigma}(h^t, q_t, r_t)$ being the expected payoff after h^t, state q_t and recommendation r_t:

$$\hat{\sigma}^1_R(\hat{h}^t) = (1-\delta)\hat{\rho}^1(\hat{h}^t) + \delta\hat{\sigma}(\hat{h}^t, q_t = 1, r_t = 1)$$

and

$$\hat{\sigma}^1_P(\hat{h}^t) = -(1-\delta)\hat{\rho}^1(\hat{h}^t) + \delta\hat{\sigma}(\hat{h}^t, q_t = 0, r_t = 1)$$

Remember that given our assumption that $\lambda < \lambda(p, q)$, for any $\hat{\rho}^1(h^t) > \hat{\rho}^0(h^t)$, the period t (not considering the future $t' > t$ payoffs) expected payoff of the agent after observing $\eta_t = 1$ is greater by reporting $r_t = 1$. This must mean that either $\hat{\sigma}(\hat{h}^t, q_t = 1, r_t = 1) < \hat{\sigma}(\hat{h}^t, q_t =$

[4] Remember that right at t itself, the principal can only follow the recommendation with some probability $\hat{\rho}(\hat{h}^t)$, without conditioning on state, as he does not know about it before funding the project, so payoffs can only differ due to some promise to follow the recommendations with different probabilities. Remember also that the actions of the agent do not change the probabilities of having a different signal realization.

$1, r_t = 0)$ or $\hat{\sigma}(\hat{h}^t, q_t = 1, r_t = 1) < \hat{\sigma}(\hat{h}^t, q_t = 1, r_t = 0)$. But then it must be the case that the principal can follow recommendation $r_{t'} = 0$ with higher probability at some history after $\hat{h}^t$ with $q_t = 1$ and increase either $\sigma_P^1(\hat{h}^t)$ or $\sigma_R^1(\hat{h}^t)$, and therefore keep the truth-telling constraint holding.

But in this case, the principal can pick a mechanism inducing increased values for either $\sigma_P^1(\hat{h}^t)$ and $\sigma_R^1(\hat{h}^t)$, by promising to either follow their advice to fund in the future or at time t with higher probability ($\tilde{\rho}^1(\bar{h}^{t'}) > \hat{\rho}^1(\bar{h}^{t'})$ for some history $\bar{h}^{t'} \in \mathcal{H}(\hat{h}^t)$ and $t' \geq t$) or not to fund with lower probability. Note that this can be done because a binding truth-telling constraint after $\eta_t = 1$ implies a non-binding constraint after $\eta_t = 0$[5] and the fact that this must mean that the principal is not following the truthful advice always if the payoff from telling the truth after a report $r_t = 0 = \eta_t$ is strictly higher than the corresponding truth-telling payoff after $\eta_t = 1$ (as the agent is biased towards $r_t = 1$ if expected future payoffs after t were the same, the continuation payoffs from reporting $r_t = 1$ would have to be higher).

Consider now the possibility that $\hat{\sigma}_R^0(\hat{h}^t) > \hat{\sigma}_P^0(\hat{h}^t)$ because for some history $h'^{t'} \in \mathcal{H}(\hat{h}^t)$ with $t' > t$, the principal is following $r_{t'} = 0$ with lower probability in some period t' and history after t succeeding a correct recommendation $r_t = 0$. The principal can then implement $r_{t'} = 0$ more often, decreasing $\tilde{\sigma}_R^0(\hat{h}^t)$ to some value in the interval $(\hat{\sigma}_P^0(\hat{h}^t), \hat{\sigma}_R^0(\hat{h}^t))$. The truth-telling constraint after $\eta_t = 1$ will be more easily satisfied and this change is desirable for the principal, as he is following recommendations matching the signal realizations more often now. The only case in which the principal may not be able to find such $\tilde{\sigma}_R^0(\hat{h}^t)$ is if the truth-telling constraint after $\hat{h}^t$ and realization $\eta_t = 0$ binds for $\hat{d}$. But then the principal can increase the probability of following $r_{t'} = 0$ after a wrong report $r_t = 1$ and decrease $\hat{\sigma}_P^1(\hat{h}^t)$ by the same amount as $\hat{\sigma}_R^0(\hat{h}^t)$ and therefore make the truth-telling constraint still hold. ■

[5]Remember that $\tilde{\sigma}_R^0(\hat{h}^t) > \tilde{\sigma}_P^0(\hat{h}^t) \geq 0$ by feasibility and by assumption.

If both truth-telling constraints are not binding, then the principal can make himself better off by following some type of recommendation more often and keeping truth-telling constraints holding, so any optimal mechanism must make one constraint bind, from the proof of Lemma 2. It is worth noting that one binding constraint implies that the other must not bind, if $p, q \in (0, 1)$, as assumed.

Lemma 3 *There is an optimal mechanism in which the truth-telling constraint after $\eta_t = 0$ binds*

Proof. As noted above, they cannot both be non-binding.

Suppose that for an optimal mechanism $\hat{d}$ we instead have the truth-telling constraint after $\eta_t = 1$ after some $\hat{h}^t$. It must be the case, given the agent's bias towards funding projects of any type, that future payoffs are lower after a report to fund. The principal can either increase the agent's payoffs after a report $r_t = 1$ (either $\sigma_R^1(\hat{h}^t)$ or $\sigma_P^1(\hat{h}^t)$) by promising to follow recommendations to fund with higher probability or decrease future payoffs after a report not to fund (either $\sigma_R^0(\hat{h}^t)$ or $\sigma_P^0(\hat{h}^t)$) by increasing the probability to follow recommendations not to fund. As this leads to a higher payoff for the principal, $\hat{d}$ cannot be optimal. ∎

We also have that there is an optimal mechanism making payoffs from $t+1$ forward equal for a report $r_t = 0$ or a correct report $r_t = 1$:

Lemma 4 *There is an optimal mechanism $\tilde{d}$ inducing $\tilde{\sigma}_R^1 = (1-\delta)\tilde{\rho}^1(h^t) + \delta\sigma^0(h^t)$ for any history h^t for some probability $\hat{\rho}^1(\hat{h}^t)$ of choosing $d_t = 1$ after $r_t = 1$.*

Proof. Suppose first that for a truthful optimal mechanism $\hat{d}$ with the truth-telling constraint after $\eta_t = 0$ binding and $\tilde{d}$ inducing $\tilde{\sigma}_R^0(h^t) = \tilde{\sigma}_P^0(h^t)$ after any history h^t it is the case that after some history $\hat{h}^t$, $\hat{\sigma}_R^1 > (1-\delta)\hat{\rho}^1(\hat{h}^t) + \delta\sigma^0(\hat{h}^t)$ for any probability $\hat{\rho}^1(\hat{h}^t)$ of picking $d_t = 1$ when $r_t = 1$. By feasibility, it must be following recommendations more often after a correct recommendation $r_t = 1$ than after $r_t = 0$, as otherwise it can be constructed through some function $\hat{\rho}^1(\hat{h}^t)$. But then the principal can always follow the (truthful) recommendations more often after a mistaken recommendation $r_t = 0$ and make the truth-telling constraint after $\eta_t = 0$ still hold (the left-hand side will only increase). The truth-constraint after $\eta_t = 1$ must also hold for a small enough increase in $\hat{\sigma}^0 + p(\hat{h}^t)$, as this constraint is not binding by assumption. ■

Finally, we conclude that AUNT with appropriate ϵ is optimal, as it maximizes the probability of following the agent's recommendations given the truth-telling constraints. To see this more clearly, note that the lemmas above show that there is a truthful mechanism, that does not punish bad recommendations not to fund a project and that treats good reports to fund in the same way as reports not to fund, and with the truth-telling constraint after observing $\eta_t = 0$ holding with equality. Take any history $\hat{h}^t$. For any $\sigma_P^1(\hat{h}^t) > 0$, the best that can be done to follow the advice and keep the truth-telling constraint after $\eta_t = 0$ holding is to follow the advice with probability 1. With the ϵ chance of getting to 0 payoff forever, we can then make the truth-telling constraint in question bind, and we are done. Remember that this is only true because UNIT is guaranteed to be strong enough to warrant truth-telling, from the condition that $\delta \geq 1/(q+1)$

5.1 Characteristics of AUNT Mechanisms

AUNT Mechanisms have some desirable features. For example, they are simple and Markovian, having the same rules at each point in time.

Note also that they do not depend on the principal's ability to observe the quality of projects that were not funded. This is a reassuring characteristic, as we can bypass the dubious assumption that even unfunded projects can have their quality revealed.

Suppose the agents have a bias towards not funding projects regardless of quality. In that case, these mechanisms can be changed straightforwardly: after a misreport of telling that a good project is bad, fund projects forever with some probability. This way, all the incentives are reversed in the same faction as before.

6 Conclusion

In many situations, expert advice is needed for decision-making in long-term repeated interactions. Not always transfers can be made to incentivize truth-telling, which might not occur because of biases held by the advisors. However, if they intrinsically care about the choices made by the decision-maker, the latter can use her commitment power to extract information.

We present a model in which one expert sees imperfect signals of the state of the world over time and a principal can only commit to decision rules as a tool to get the signal information. We find that a simple mechanism is optimal for the principal, generalizing mechanisms ignoring the expert's advice for a number of periods, for example. The mechanism is helpful as a first step in exploring what is the frontier of what can be used in settings such as the one considered here.

The main intuition for the main results of the paper is that the agent with a state-independent utility is insured against bad realizations, therefore, has more interest in keeping being listened to. The principal can explore this gap and extract a lot of the informational surplus from the agent. More general structures keeping these features should have similar results.

Many venues for future work are left to be explored. The more salient is trying to go around the assumption that bias is known but that it might instead be learned through the mechanism. Another interesting venue would be to assume that precision is unknown and learning can happen over time. Deb et al. (2018) present a setting in this vein, where the principal must hire the most precise expert, and has to choose a perfect hiring time. A more straightforward extension for the setting presented here would be to keep the infinite horizon and make punishments for misreporting depend on information on past accuracy. A grim-trigger mechanism, such as AUNT, could no longer be optimal, given the perspective of future gains from information coming from knowing the precision better.

References

Azoulay, P. and Li, D. (2020). Scientific grant funding. Working Paper 26889, National Bureau of Economic Research.

Azoulay, P., Zivin, J. S. G., and Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3):527–554.

Balsmeier, B., Fleming, L., and Manso, G. (2017). Independent boards and innovation. *Journal of Financial Economics*, 123(3):536–557.

Bhattacharya, J. and Packalen, M. (2020). Stagnation and scientific incentives. Working Paper 26752, National Bureau of Economic Research.

Bhattacharya, S. and Mukherjee, A. (2013). Strategic information revelation when experts compete to influence. *The RAND Journal of Economics*, 44(3):522–544.

Che, Y.-K. and Kartik, N. (2009). Opinions as incentives. *Journal of Political Economy*, 117(5):815–860.

De Clippel, G., Eliaz, K., Fershtman, D., and Rozen, K. (2019). On selecting the right agent. *Working Paper*.

Deb, R., Pai, M. M., and Said, M. (2018). Evaluating strategic forecasters. *American Economic Review*, 108(10):3057–3103.

Gerardi, D., McLean, R., and Postlewaite, A. (2009). Aggregation of expert opinions. *Games and Economic Behavior*, 65(2):339–371.

Gerardi, D. and Yariv, L. (2008). Costly expertise. *American Economic Review*, 98(2):187–93.

Glazer, J. and Rubinstein, A. (1998). Motives and implementation: On the design of mechanisms to elicit opinions. *Journal of Economic Theory*, 79(2):157–173.

Guo, Y. and Hörner, J. (2018). Dynamic allocation without money. *Northwestern University and Yale University*.

Jackson, M. O. and Sonnenschein, H. F. (2007). Overcoming incentive constraints by linking decisions 1. *Econometrica*, 75(1):241–257.

Li, D. (2017). Expertise versus bias in evaluation: Evidence from the nih. *American Economic Journal: Applied Economics*, 9(2):60–92.

Li, J., Matouschek, N., and Powell, M. (2017). Power dynamics in organizations. *American Economic Journal: Microeconomics*, 9(1):217–41.

Lipnowski, E. and Ramos, J. (2019). Repeated delegation. *Available at SSRN 2552926*.

Manso, G. (2011). Motivating innovation. *The Journal of Finance*, 66(5):1823–1860.

Margaria, C. and Smolin, A. (2018). Dynamic communication with biased senders. *Games and Economic Behavior*, 110:330–339.

Meng, D. (2018). Optimal mechanisms for repeated communication (job market paper). *Unpublished Manuscript*.

Pakzad-Hurson, B. (2021). Crowdsourcing and optimal market design. *Available at SSRN 2618837*.

Stephan, P. E. (2012). *How economics shapes science*, volume 1. Harvard University Press Cambridge, MA.

Appendix

Almost-perfect Precision

The next theorem shows that the principal can get arbitrarily close to the first-best when p is close enough to 1 and the agent is patient enough. Therefore, when precision is almost-perfect, the principal can get very close to the first-best.

Theorem 3 *Take any $\epsilon > 0$. If $\delta > 1/(1+q)$, $\exists \gamma \in (0,1)$ such that if $p \geq 1-\gamma$, the UNIT mechanism is such that $U^P \geq U^{FB} - \epsilon$.*

Proof. We will use the UNIT mechanism as a mechanism as required. Let's use the terms σ_R as the "reward" continuation payoff for when a report to fund a project is not shown to be a misreport of the signal and σ_P as the continuation payoff for when this misreport is detected. From the mechanism itself, $\sigma_P = 0$. Given a signal $\eta_t = 0$, the agent will tell the truth if:

$$\delta\sigma_R \geq (1-\delta)(\lambda(Pr(q_t = 1|\eta_t = 0) - Pr(q_t = 0|\eta_t = 0)) + (1-\lambda) + \delta(Pr(q_t = 1|\eta_t = 0)\sigma_R)$$

or

$$\delta\sigma_R \geq (1-\delta)\left(\lambda\left(\frac{(1-p)q}{(1-p)q+(1-q)p} - \frac{(1-q)p}{(1-q)p+(1-p)q}\right) + (1-\lambda)\right) + \delta\frac{(1-p)q}{(1-p)q+(1-q)p}\sigma_R$$

leading to

$$\sigma_R \geq \frac{(1-\delta)[(1-p)q + (1-2\lambda)p(1-q)]}{\delta(1-q)p}$$

The promise-keeping constraint gives us:

$$\sigma_R = \frac{(1-\delta)(pq + (1-2\lambda)(1-q)(1-p))}{1-\delta(p+q-pq)}$$

This equation is telling us that given truth-telling, this is what the agent can get as the expected payoff, given the prior on the future project qualities and taking into account that the principal will commit to the mechanism structure chosen at period $t = 0$.

Using these two equations, we get the condition:

$$\frac{pq + (1-2\lambda)(1-q)(1-p)}{1-\delta(p+q-pq)} \geq \frac{q + (1-2\lambda)p - 2pq(1-\lambda)}{\delta(1-q)p}$$

This leads us to a condition on δ, with $\kappa = pq + (1-q)(1-p)$:

$$\delta \geq \frac{1-\kappa-2\lambda p(1-q)}{(1-q)p[\kappa - 2\lambda(1-q)(1-p)] + (p+q-pq)(1-\kappa-2\lambda p(1-q))}$$

The right-hand side is decreasing in λ. So for $\delta \geq (1-\kappa)/[p(1-q)+q(1-\kappa)]$ the condition will hold for any λ.

One must also check whether this threat when a bad project is funded is not so strong that an agent seeing $\eta_t = 1$ would not prefer reporting $r_t = 0$, given the possibility of the signal being wrong about the state of the world. We need:

$$\sigma_R \leq (1-\delta)\frac{pq + (1-2\lambda)(1-p)(1-q)}{1-\delta(p+q-pq)}$$

By promise-keeping, we already know that σ_R is given by:

$$\sigma_R = \frac{(1-\delta)(pq + (1-2\lambda)(1-q)(1-p))}{1-\delta(p+q-pq)}$$

For any δ, then, we have that the inequality above holds, so this condition will be satisfied here.

The principal's utility will be equal to:

$$U^P(d^{UM}) = \frac{(1-\delta)(p+q-1)}{1-\delta+\delta(1-q)(1-p)}$$

To understand this expression, note that as the agent is kept honest, there is a probability $1-(1-p)(1-q)$ that the payoffs continue and a complementary probability of getting payoff 0 forever after. The first-best principal payoffs are given by $p+q-1$. Therefore when $p \to 1$, we get that the payoffs get arbitrarily close to it, at the value q, as desired. ■