Robust contracting and voluntary disclosure*

Eran Hanany† Peter Klibanoff‡

This Version: December 12, 2025

Abstract

This paper analyzes contracting between a principal and an agent when the principal is uncertain exactly which actions may be feasible for the agent and has a strong desire for robustness (in the worst-case or maxmin sense) of the expected profits generated. A prominent and path-breaking paper in this direction is Carroll (2015), which demonstrates that linear contracts are robustly (worst-case) optimal given uncertainty about an agent's available actions. What if, when it is in their interest, the agent could choose to disclose that they have access to a particular additional action, and such statements could be verified by the principal? Does this change the form of robustly optimal contracts offered to an agent who either chooses not to disclose or has no additional action to disclose? Are such contracts still linear? We show that voluntary disclosure can substantially change the form of robustly optimal contracts. In particular, we show the possibility of and provide sufficient conditions for equilibrium contracts offered following non-disclosure to be non-linear. This equilibrium non-linearity does not always occur. We show that linearity results when there are few publicly known-to-be-available actions that generate a positive surplus.

1 Introduction

This paper analyzes contracting between a principal and an agent when the principal is uncertain exactly which actions may be feasible for the agent and has a strong desire for ro-

*We thank Nemanja Antic, Ashwin Kambhampati and audiences at the Technion and Tel Aviv University game theory seminars, the IMSI 2022 workshop "Decision Making Under Uncertainty" and the ORSIS 2023, RUD 2024 and D-TEA 2025 conferences for helpful discussion. This research was partially supported by Grant No. 2022020 to Hanany and Klibanoff from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. The authors have no other financial or other material interests related to this research to disclose.

†Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel. Email: hananye@post.tau.ac.il

‡Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. E-mail: peterk@kellogg.northwestern.edu

bustness (in the worst-case or maxmin sense) of the expected profits generated. A prominent and path-breaking paper in this direction is Carroll (2015), which demonstrates that linear contracts are robustly (worst-case) optimal given uncertainty about an agent's available actions. Equivalently, in language more familiar in decision theory, linear contracts are shown to be optimal for a Maxmin Expected Utility (MEU), risk-neutral principal when facing complete ambiguity or imprecision regarding the actions (if any) the agent may privately have available (in addition to some publicly known available actions).

In this paper, we address the following question: What if the agent could choose to disclose that they have access to a particular additional action, and such statements could be verified by the principal? Does this change the form of robustly optimal contracts offered to an agent who either chooses not to disclose or has no additional action to disclose? Are such contracts still linear?

Why consider voluntary disclosure and uncertainty about the agent's available actions? A prominent example we have in mind is a venture capitalist (the principal, henceforth VC in this example) contracting with an entrepreneur (the agent). In many entrepreneurial start-ups, a key feature of the financial contracting environment is private information of the entrepreneur about the technologies they have available or hope to develop. At the same time, entrepreneurs often try to improve their contracting position through extensive voluntary disclosure of the details of these possibilities to the VC. Importantly, while such disclosures may be largely able to be vetted and verified by the VC once presented, the very nature of entrepreneurial novelty and insight suggests that the VC would have trouble distinguishing between an entrepreneur with nothing to disclose beyond what the VC already knows about their business, and an entrepreneur who intentionally withholds some information while claiming that they have nothing more to disclose.

Our formal model is presented in Section 2, including the payoffs and objectives of the principal and agent, the definitions of contracts and agent's actions, the structure and timing of the contracting game including an initial voluntary disclosure stage, and the definition of an equilibrium of this game. Section 3 contain two sets of results. Theorems 3.1 and 3.3 show that voluntary disclosure can substantially change the form of robustly optimal contracts. In particular, the first of these results shows the possibility that all equilibrium contracts offered following non-disclosure are non-linear, in contrast to the linear equilibrium contracts identified by Carroll (2015) in a model without disclosure. The second of these results provides sufficient conditions for there to be a non-linear contract offered in equilibrium following non-disclosure. This equilibrium non-linearity does not always occur. Theorem 3.4 provides sufficient conditions for there to be a linear contract offered in equilibrium following non-disclosure. In particular, we show that linearity results when there are few publicly known-to-be-available actions that generate a positive surplus.

Before turning to the model, we discuss the relation of our paper to several strands of the literature. The first is the huge literature on Principal-Agent models and the associated contracting problem – especially papers examining the problem of contracting with a risk-neutral agent under limited liability constraints, prominent examples being Innes (1990) and Diamond (1998). An influential view/critique of the approaches taken in much of the extant literature on contracting, pricing and mechanism design is sometimes called the Wilson critique (attributed to Robert Wilson) and described in Carroll (2019) as advocating that "realistic mechanisms should not be finely tuned to parametric assumptions, such as probability distributions of values or functional forms of preferences". Motivated in part by this view, a recent direction in the literature is to investigate contracts that are optimal in a robust sense with respect to some such assumptions. See Carroll (2019) for a recent survey of literature on robust contracting and mechanism design and further discussion of this and similar motivations.

The key innovation in our paper is combining robust contracting with the option of verifiable voluntary disclosure. Though we believe we are the first to bring together robust contracting with such voluntary disclosure, our modeling of this disclosure follows literature adopting the idea that while a given piece of evidence may be verifiably disclosed, it is not possible to verify the absence of such evidence. Such an information structure was first described and modeled in Dye (1985). Models with such disclosure or partial disclosure possibilities have been analyzed in various settings including (quoting from Esö and Wallace (2022, p.4)) "Shin (1994b) in a pure communication game featuring two senders with opposing interests; Shin (1994a, 2003) in an exchange economy and an asset-pricing model; Shavell (1994) in a bilateral trading environment; Glode, Opp, and Zhang (2018) in a model with screening; Acharya, DeMarzo, and Kremer (2011) as well as DeMarzo, Kremer, and Skrzypacz (2019) in asset markets (which they reinterpret for other applications as well); and Esö and Wallace (2014, 2019) in dynamic bargaining games." This literature also includes papers generalizing beyond the Dye (1985) structure, including, among others, Ben-Porath and Lipman (2012) on implementation with partial provability and Ben-Porath, Dekel and Lipman (2019) on mechanism design with evidence. The only principal-agent analyses with evidence that we are aware of are in the original Dye (1985, Section 3) paper and in Gode and Singh (2006). Neither focuses beyond adverse selection nor includes concern for robustness.

Carroll (2015) has led to an active literature on robust principal-agent contracting with action uncertainty. Antić (2021), Walton and Carroll (2022) and Olszewski (2025) explore conditions on the action uncertainty that either are sufficient for linearity of robustly optimal contracts or that lead to their non-linearity. Antić (2021) shows that when the possible actions are bounded below by a not-so-bad moderate technology, robustly optimal contracts may be non-linear – and specifically a combination of debt and linear (i.e., debt and equity). Walton and Carroll (2022) show that an important condition in generating linearity of robustly optimal contracts is a particular type of richness of the action uncertainty, together with an assumption about the agent's responsiveness to incentives. Olszewski (2025) shows that the family of action uncertainties under which Carroll's linearity result holds is small in a topological sense. Burkett and Rosenthal (2024) and Antić and Georgiadis (2025) modify the Carroll (2015) setting by assuming that the only restriction on the action uncertainty comes in the form of a finite set of observed (contract, output distribution) pairs, with the interpretation that at least some actions having the output distributions appearing in the pairs are available to the agent, and each output distribution is part of an action that is a best response of the agent to the corresponding contract. They investigate robustly optimal contracts and show that in many cases they take the form of a convex combination of the observed contracts and a linear contract. Like all of these papers and Carroll (2015), we focus on pure strategy contract offers. Kambhampati (2023) shows that the principal in Carroll (2015) can do better by randomizing over linear contracts, and Kambhampati et al. (2025) characterizes the robustly optimal contract when randomization is allowed and shows that it is a randomization over linear contracts. In Theorem 3.5 we show that a robustly optimal contract when randomization is allowed is robust to voluntary disclosure in the sense that it remains an equilibrium contract following non-disclosure in the game where voluntary disclosure is allowed.

Observe that the equilibrium voluntary disclosure in our paper makes the action uncertainty faced by the principal a function of their equilibrium contract offers following disclosure and non-disclosure. Specifically, in equilibrium, if the agent, anticipating these contracts, does not make a disclosure, the principal can rule out the availability of those actions that would have been disclosed. This contrasts with the exogenous action uncertainty in the literature. Thus, whether or not linear contracts are robustly optimal when disclosure is allowed does not follow from existing results.

2 Model

The formal model builds on Carroll (2015), with the main change being an initial disclosure opportunity for the agent. There are two players, a principal and an agent. The principal is uncertain about exactly which output-generating actions the agent has available. The structure of the game between them is the following: first, the agent chooses between verifiably disclosing the actions they have available and providing no disclosure; second, having observed any disclosure or non-disclosure, the principal offers the agent a wage contract mapping output to payments; third, the agent chooses from among their available actions; finally, output is realized and payments are made according to the wage contract.

In more detail, the set of possible output levels $\mathcal{Y}$ is a compact subset of the non-negative reals with minimum normalized to zero. This ensures that $\Delta(\mathcal{Y})$, the set of all Borel probability distributions on $\mathcal{Y}$, is compact. An action for the agent is a pair $[q, d] \in \Delta(\mathcal{Y}) \times \mathbb{R}_+$ consisting of a probability distribution over output and a non-negative cost. The agent has a compact set of available actions, $\mathcal{A}$, which is commonly known to contain a compact set of actions, $\mathcal{A}_0$. We assume throughout, as in Carroll (2015), that there is at least one action in $\mathcal{A}_0$ such that $E_q[y] - d > 0$, where $E_q[\cdot]$ is the expectation operator with respect to q. Thus, it is common knowledge that a surplus generating action is available. In addition to the actions in $\mathcal{A}_0$, the agent may have available some additional action, $[p, c] \in \Delta(\mathcal{Y}) \times \mathbb{R}_+$, and the identity of this action is known only to the agent. Note that the special case where the agent has only the actions in $\mathcal{A}_0$ available is modelled by $[p, c] \in \mathcal{A}_0$.[1] After observing any disclosure or non-disclosure, the principal offers the agent a continuous wage contract $w : \mathcal{Y} \to \mathbb{R}_+$. While Carroll assumes limited-liability on the part of the agent (justifying the non-negativity of wages), we will additionally assume limited liability on the part of the principal, implying that wage contracts must satisfy $w(y) \leq y$ for each $y \in \mathcal{Y}$. Denote the set of all such contracts by $\mathbb{W}$.

Since the agent's set of available third-stage actions is $\mathcal{A} = \mathcal{A}_0 \bigcup \{[p, c]\}$, we will generally denote the agent by their additional action, $[p, c]$. At the first stage, agent $[p, c] \notin \mathcal{A}_0$ chooses between the verifiable disclosure $[p, c]$ and non-disclosure, which we denote by $\emptyset$. An agent $[p, c] \in \mathcal{A}_0$ has only the choice of non-disclosure. This restriction of available choices as a function of the additional action $[p, c]$ is as in the strand of literature on voluntary disclosure of verifiable evidence starting from Dye (1985). The idea is that an agent claiming to have a certain action available can, if required, prove that to the principal, thereby preventing false disclosures. However, the agent has no way to prove that they do *not* have access to any additional action beyond those in $\mathcal{A}_0$, and thus an agent with $[p, c] \in \mathcal{A}_0$ cannot credibly disclose this and must choose non-disclosure. This completes the description of the actions and information available to the principal and agent at each point in the game.

A strategy profile for this game is a vector $(\mathcal{N}, \mathbf{w}, A)$, where $\mathcal{N} \supseteq \mathcal{A}_0$ is the set of agents $[p, c]$ not disclosing their additional action, $\mathbf{w} = (w^{\emptyset}, (w^{[p,c]})_{[p,c] \in (\Delta(\mathcal{Y}) \times \mathbb{R}_+) \setminus \mathcal{A}_0})$ are the contracts offered by the principal following non-disclosure and each possible disclosure,

[1] Carroll (2015) allows the agent to potentially have multiple additional actions available simultaneously. However, he shows in his Section II.B that his results go through when the agent can have at most one additional action. What is important for his results is the richness of what the additional action might be, which we maintain.

respectively, and $A : (\Delta(\mathcal{Y}) \times \mathbb{R}_+) \times \{0,1\} \times \mathbb{W} \to \Delta(\mathcal{Y}) \times \mathbb{R}_+$ such that $A([p,c],\varrho,w) \in \mathcal{A}_0 \bigcup \{[p,c]\}$ is the action taken at the third stage by agent $[p,c]$ given disclosure decision $\varrho = 1$ if $[p,c]$ was disclosed and $\varrho = 0$ otherwise and given that contract w was offered by the principal. When we refer to the output distribution and cost components of $A([p,c],\varrho,w)$ separately, we denote them by $A_1([p,c],\varrho,w)$ and $A_2([p,c],\varrho,w)$, respectively.

2.1 Definition of Equilibrium

The agent maximizes expected wages net of action cost. Thus, given available actions $\mathcal{B}$ and a contract w, the agent obtains expected payoff:

Definition 2.1 (Agent's expected payoff given some compact set of available actions $\mathcal{B}$)**.**

$$V_A(w \mid \mathcal{B}) = \max_{[q,d]\in\mathcal{B}} E_q[w(y)] - d$$

The global tie-breaking assumption used by Carroll (2015) and adopted by us is that whenever an agent has multiple actions in $\mathcal{A}$ generating $V_A(w \mid \mathcal{A})$, they break the indifference in favor of what the principal prefers. The principal's payoff is expected profit, namely expected output minus expected wage. Formally, an agent $[p,c]$'s set of optimal actions given w is:

Definition 2.2 (Agent's set of optimal actions given w, the agent's technology and ties broken in favor of the principal)**.**

$$\mathcal{A}^*(w \mid [p,c]) = \arg \max_{[q,d]\in\arg\max_{[r,e]\in\mathcal{A}_0 \bigcup\{[p,c]\}} E_r[w(y)]-e} E_q[y - w(y)]$$

The principal's payoff when offering a contract w and facing an agent $[p,c]$ who best responds to w is:

Definition 2.3 (Principal's payoff given w and the agent's technology)**.**

$$V_P(w \mid [p,c]) = E_q[y - w(y)] \text{ for } [q,d] \in \mathcal{A}^*(w \mid [p,c])$$

In the special case where $[p,c] \in \mathcal{A}_0$, we denote this payoff by $V_P(w \mid \mathcal{A}_0)$.

Since we consider disclosure, the principal chooses not one contract, but rather $\mathbf{w}$, specifying a (possibly different) contract following each possible disclosure or non-disclosure. Following a disclosure $[p,c]$, it is clear that the principal chooses the contract w to maximize $V_P(w \mid [p,c])$. It is less obvious what the principal should maximize following non-disclosure. In line with Carroll (2015) and a variety of literature related to robust optimization, we assume that when the principal is uncertain about which agent they are facing, they value a contract by its worst case guaranteed expected profit. This is also related to extreme versions of the maxmin expected utility (MEU) criterion in the literature on ambiguity aversion following Gilboa and Schmeidler (1989) or imprecision aversion following Gajdos et al. (2008). Formally, when the principal knows only that the agent $[p, c]$ they are facing lies in some set $\mathcal{C} \subseteq \Delta(\mathcal{Y}) \times \mathbb{R}_+$, the principal is assumed to maximize the worst case expected profit, where the worst case is taken over $[p, c] \in \mathcal{C}$. Thus this worst-case expected profit when offering contract w and with no restrictions on $[p, c]$ is:

Definition 2.4 (Principal's worst-case payoff when offering w ignoring disclosure)**.**

$$V_P(w) = \inf_{[p,c]} V_P(w \mid [p, c])$$

In the Carroll (2015) model, $V_P(w)$ is what the principal maximizes when choosing w.[2] Since we consider disclosure, following non-disclosure, given $\mathcal{N}$, the principal infers that $[p, c] \in \mathcal{N}$. Therefore the principal maximizes this constrained worst-case following non-disclosure:

Definition 2.5 (Principal's worst-case payoff following non-disclosure when offering w, given non-disclosure set $\mathcal{N}$)**.**

$$V_P^D(w; \mathcal{N}) = \inf_{[p,c] \in \mathcal{N}} V_P(w \mid [p, c])$$

We use the definitions above to define an equilibrium as follows:

Definition 2.6 (Equilibrium)**.** A strategy profile $(\mathcal{N}, \mathbf{w}, A)$ is an equilibrium if:

(i) for all $([p, c], \varrho, w)$,

$$A([p, c], \varrho, w) \in \mathcal{A}^*(w \mid [p, c]),$$

(ii) for all $[p, c] \notin \mathcal{A}_0$,

$$[p, c] \notin \mathcal{N} \text{ implies } V_A(w^{[p,c]} \mid \mathcal{A}_0 \bigcup \{[p, c]\}) \geq V_A(w^{\emptyset} \mid \mathcal{A}_0 \bigcup \{[p, c]\}) \text{ and}$$
$$[p, c] \in \mathcal{N} \text{ implies } V_A(w^{\emptyset} \mid \mathcal{A}_0 \bigcup \{[p, c]\}) \geq V_A(w^{[p,c]} \mid \mathcal{A}_0 \bigcup \{[p, c]\}),$$

and

(iii) for all $[p, c] \notin \mathcal{A}_0$,

$$w^{[p,c]} \in \arg\max_{w \in \mathbb{W}} V_P(w \mid [p, c]),$$

[2] As mentioned in the Introduction, this restriction to pure strategy contract offers is not without loss of generality (Kambhampati (2023)). See Section 3.4 for consideration of randomized contracts.

and

$$w^{\emptyset} \in \arg\max_{w \in \mathbb{W}} V_P^D(w; \mathcal{N}).$$

Condition (i) says that the agent best responds at the third stage to any contract offer, with ties broken in favor of the principal. Condition (ii) says that any agent who has something to disclose makes the disclosure decision optimally given the contract offers and agent actions specified in the strategy profile. Finally, condition (iii) says that in each disclosure/non-disclosure contingency, the principal's contract offer is a (maxmin) best response given the disclosure and action strategies.[3,4] Finally, when the contract $w^{\emptyset}$ is offered following non-disclosure, and $\mathcal{N}(w^{\emptyset})$ is a non-disclosure set satisfying the best response condition (ii) in Definition 2.6, we denote the principal's worst-case payoff by:

Definition 2.7 (Principal's worst-case payoff when offering w following non-disclosure given a strategy profile specifying that $w^{\emptyset}$ should be offered)**.**

$$V_P^D(w) \equiv V_P^D(w; \mathcal{N}(w^{\emptyset})).$$

Denote by $w^{[p,c],[q,d]}$ a profit-maximizing contract among those implementing action $[q, d]$ from the set $\mathcal{A}_0 \bigcup \{[p, c]\}$. Note that condition (iii) says that $V_P(w^{[p,c]} \mid [p, c]) = \max_{[q,d] \in \mathcal{A}_0 \bigcup \{[p,c]\}} V_P(w^{[p,c],[q,d]} \mid [p, c])$.

3 Results

3.1 Voluntary Disclosure can lead to non-linear robustly optimal contracts

The result stated and proved below shows, by example, that voluntary disclosure can lead to non-linearity of all equilibrium robustly optimal contracts following non-disclosure, in contrast to the linearity of robustly optimal contracts when no disclosure is allowed. The

[3] An additional condition that is arguably desirable for equilibrium is ex-ante best response on the part of the principal and agent (this would be needed, for example, if Sequential Optimality (Hanany, Klibanoff and Mukerji, 2020) is taken as the equilibrium concept). Ex-ante best response follows directly from conditions (i) and (ii) for the agent since their best response is separable across information sets. For the principal, the ex-ante best response condition is $\mathbf{w} \in \arg\max_{\mathbf{w}} \inf_{[p,c]} \{V_P(w^{\emptyset} \mid [p, c])\mathbb{I}_{[p,c] \in \mathcal{N}} + V_P(w^{[p,c]} \mid [p, c])\mathbb{I}_{[p,c] \notin \mathcal{N}}\}$. A nice feature of this worst-case contracting setting is that this additional condition follows from condition (iii) since the principal's ex-ante worst-case set is taken to be the set of all (distributions over) additional actions $[p, c]$, and that this is updated after disclosure to be the $[p, c]$ that is disclosed, and after non-disclosure to be the set of all (distributions over) $[p, c] \in \mathcal{N}$.

[4] Since the agent is optimizing with ties broken in favor of the principal, the principal's payoff as formulated in Condition (iii) is correct in that it is always consistent with the specific action the agent is taking according to the strategy profile.

intuition for this finding is simple at the broadest level – with enough disclosure, the richness of the set of additional actions which are not disclosed, and thus over which uncertainty remains, may not be sufficient to work against all non-linearities in the contract. However, because the decision to disclose or not is an equilibrium choice of the agent, it is not at all obvious whether and in what circumstances equilibrium disclosure reduces the uncertainty about additional actions in a way that supports non-linearity of the robustly optimal contract offered following nondisclosure. Inspection of the proof below is an illustration that the arguments involved in showing this can be complex. They involve analysis of solutions and/or bounds on solutions of non-trivial LPs (linear programs), for example.

Theorem 3.1. *There exists a commonly known technology $\mathcal{A}_0$ such that, (1) when no disclosure is allowed, the unique equilibrium robust contract is a positive linear contract, and (2) when voluntary disclosure is allowed, there is an equilibrium in which the robust contract offered following non-disclosure is a non-linear contract, and (3) when voluntary disclosure is allowed, there is no equilibrium in which the robust contract offered following non-disclosure is a linear contract.*

Proof of Theorem 3.1. The proof will be by example. Let the set of possible outputs $\mathcal{Y} = \{0, 1, 2\}$ and the commonly known technology

$$\mathcal{A}_0 = \{[(1,0,0),0], [(\frac{22}{32}, \frac{1}{32}, \frac{9}{32}), \frac{3}{200}], [(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}], [(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]\}.$$

Suppose first that no disclosure is allowed. By Carroll (2015), since all positive expected output actions in $\mathcal{A}_0$ have full support on $\mathcal{Y}$, any equilibrium robust contract is linear. Furthermore, such linear contracts have coefficient $\beta = \sqrt{\frac{d}{E_q[y]}}$ for $[q, d] \in \arg\max_{[r,e]\in\mathcal{A}_0} \sqrt{E_r[y]} - \sqrt{e}$. For this example, the unique such $\beta = \frac{1}{5}$. This establishes (1).

Now consider the game where disclosure is allowed. Since disclosure can only remove some additional actions from the principal's consideration following non-disclosure, allowing disclosure can never strictly lower the worst-case payoff of a contract following non-disclosure, meaning $V_P(w) \leq V_P^D(w)$ for any contract w. Furthermore, since an agent with only the actions in $\mathcal{A}_0$ available can never disclose, an upper bound on $V_P^D(w)$ is given by the principal's payoff from offering w to an agent having only $\mathcal{A}_0$ available. Thus $V_P^D(w) \leq V_P(w \mid \mathcal{A}_0)$.

We next show that when voluntary disclosure is allowed, it is part of an equilibrium for the non-linear contract $w = (0, \frac{3}{25}, 0)$ to be offered following non-disclosure. Observe that for an agent having only $\mathcal{A}_0$ available, the equilibrium response to the contract w is $\mathcal{A}^*(w \mid \mathcal{A}_0) = \{[(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}]\}$. Thus, $V_P(w \mid \mathcal{A}_0) = E_{(\frac{1}{2},\frac{1}{4},\frac{1}{4})}[y] - \frac{3}{100} = \frac{72}{100}$ and therefore $V_P^D(w) \leq \frac{72}{100}$. We claim that in fact this bound is tight and $V_P^D(w) = \frac{72}{100}$. Observe that this is equivalent to showing that for all agents $[p, c]$ such that $E_p[y - w(y)] < \frac{72}{100}$ and

$E_p[w(y)] - c > V_A(w \mid \mathcal{A}_0) = 0$ (i.e., such that the agent would respond to w with action $[p, c]$ and this would yield the principal a payoff strictly below $\frac{72}{100}$), $[p, c] \notin \mathcal{N}$.

Observe that if $[p, c]$ is such that the principal would implement an action $[q, d] \in \mathcal{A}_0$ following disclosure, then the contract following disclosure, $w^{[p,c]}$, must solve the program

$$\max_{w(1),w(2)} E_q[y] - q(1)w(1) - q(2)w(2)$$

$$\text{s.t.}$$

$$\begin{aligned} & q(1)w(1) + q(2)w(2) - d \geq V_A(w \mid \mathcal{A}_0) \\ & q(1)w(1) + q(2)w(2) - d \geq p(1)w(1) + p(2)w(2) - c \qquad (3.1) \\ & w(1) \geq 0, w(2) \geq 0. \end{aligned}$$

First, suppose $[p, c]$ is such that the principal would implement the action $[q, d] = [(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ following disclosure. The solution to the program (3.1) with this $[q, d]$ has an upper bound obtained by dropping the constraint involving $[p, c]$ and solving the relaxed program. This upper bound on the payoff to the principal is $\frac{5679}{6100}$ and the corresponding lower bound on the payoff to the agent is $\frac{51}{244}$, and these are achieved with the contract $w^0 = (0, \frac{219}{1220}, \frac{4737}{6100})$. Thus, under this supposition, $V_A(w^{[p,c]} \mid [p, c]) \geq \frac{51}{244} > \frac{3}{25} \geq \frac{3}{25}p(1) - c > 0$, where the last inequality follows from $E_p[w(y)] - c > V_A(w \mid \mathcal{A}_0) = 0$, and therefore $[p, c] \notin \mathcal{N}$.

Second, suppose $[p, c]$ is such that the principal would implement the action $[p, c]$ following disclosure. This yields the principal at most $E_p[y] - c$, since $E_p[w^{[p,c]}(y)] \geq c$ as the agent can always guarantee themselves a non-negative payoff by taking action $[(1, 0, 0), 0] \in \mathcal{A}_0$. Recall that $E_p[y] < \frac{3}{25}p(1) + \frac{72}{100}$ and $\frac{3}{25}p(1) > c$. This implies that $\frac{25}{3}c < E_p[y] < \frac{9}{11}$ and $c < \frac{3}{25}\frac{9}{11} = \frac{27}{275}$. Letting $E_p[y] = k$, we can re-write $E_p[w^{[p,c]}(y)] - c$ as $kw^{[p,c]}(1) + p(2)(w^{[p,c]}(2) - 2w^{[p,c]}(1)) - c$ where $p(2) \in [\max\{0, k-1\}, \frac{k}{2}]$. Furthermore, $E_p[y] < \frac{3}{25}p(1) + \frac{72}{100}$ may be re-written as $p(2) < \frac{36}{100} - \frac{44}{100}p(1) = \frac{36}{100} - \frac{44}{100}(k - 2p(2))$ which is equivalent to $p(2) < 3 - \frac{11}{3}k$. We return to program (3.1) with $[q, d] = [(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ and tighten by replacing the right-hand side of the constraint involving $[p, c]$ with the upper bound $kw(1) + \max\{\max\{0, k-1\}(w(2) - 2w(1)), \min\{\frac{k}{2}, 3 - \frac{11}{3}k\}(w(2) - 2w(1))\} - 0$. Solving the tightened program as a function of k for $0 < k < \frac{9}{11}$ yields a lower bound on the principal's payoff following disclosure of $[p, c]$ that, when $0 < k < \frac{9}{11}$, is always strictly above $k \geq k - c$, proving that the principal does strictly better following disclosure of $[p, c]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ instead of $[p, c]$.

Third, suppose $[p, c]$ is such that the principal would implement the action $[q, d] = [(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}]$ following disclosure. The solution to the program (3.1) with this $[q, d]$ has an upper bound obtained by dropping the constraint involving $[p, c]$ and solving the relaxed program. This yields an upper bound on the payoff to the principal of $\frac{72}{100}$ and a corresponding lower bound on the payoff to the agent of 0. Since, as was established above, the principal can get at least $E_p[y]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ following disclosure of $[p, c]$, for the principal to want to implement $[(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}]$, it must be that $E_p[y] \leq \frac{72}{100}$. We can calculate, again using the principal's lower bound from implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ as a function of k established above, that adding the restriction $k \leq \frac{72}{100}$ yields a new bound, $\frac{84}{100}$, derived by taking the minimum over such k. Since $\frac{84}{100} > \frac{72}{100}$, the principal does strictly better following disclosure of $[p, c]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ instead of $[(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}]$.

Fourth, suppose $[p, c]$ is such that the principal would implement the action $[q, d] = [(\frac{22}{32}, \frac{1}{32}, \frac{9}{32}), \frac{3}{200}]$ following disclosure. The solution to the program (3.1) with this $[q, d]$ has an upper bound obtained by dropping the constraint involving $[p, c]$ and solving the relaxed program. This yields an upper bound on the payoff to the principal of $\frac{463}{800}$ and a corresponding lower bound on the payoff to the agent of 0. Since, as was established above, the principal can get at least $E_p[y]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ following disclosure of $[p, c]$, for the principal to want to implement this $[q, d]$, it must be that $E_p[y] \leq \frac{463}{800} < \frac{72}{100}$. Again using the principal's lower bound of $\frac{84}{100}$ from implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ when $k \leq \frac{72}{100}$, since $\frac{84}{100} > \frac{463}{800}$, the principal does strictly better following disclosure of $[p, c]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ instead of $[(\frac{22}{32}, \frac{1}{32}, \frac{9}{32}), \frac{3}{200}]$.

Finally, suppose $[p, c]$ is such that the principal would implement the action $[q, d] = [(1, 0, 0), 0]$ following disclosure. The principal's payoff from doing so would be 0. Since, as was established above, the principal can get at least $E_p[y]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ following disclosure of $[p, c]$, for the principal to want to implement this $[q, d]$, it must be that $E_p[y] \leq 0$. Since $[p, c]$ was assumed to satisfy $E_p[w(y)] - c > 0$, the principal does strictly better following disclosure of $[p, c]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ instead of $[(1, 0, 0), 0]$.

We have now completed the argument that for all additional actions $[p, c]$ such that $E_p[y] - E_p[w(y)] < \frac{72}{100}$ and $E_p[w(y)] - c > V_A(w \mid \mathcal{A}_0) = 0$, $[p, c] \notin \mathcal{N}$, and have therefore established that $V_P^D(w) = \frac{72}{100}$ for the non-linear contract $w = (0, \frac{3}{25}, 0)$.

It remains to show that there is no other contract w' that is a robustly strictly profitable deviation from w following non-disclosure. Formally, we want to show that for each $w' \neq w$, $V_P^D(w'; \mathcal{N}(w)) \leq \frac{72}{100} (= V_P^D(w))$, where $\mathcal{N}(w)$ is the non-disclosure set satisfying the best response condition (ii) in Definition 2.6 when $w^{\emptyset} = w$. First, observe that since the action $[(1, 0, 0), 0] \in \mathcal{A}_0$, for any contract w' and any actions $[q, d]$ and $[p, c]$, if $[q, d] \in \mathcal{A}^*(w' \mid [p, c])$ then $E_q[w'(y)] \geq d$. Therefore, $V_P^D(w'; \mathcal{N}(w)) \leq V_P(w' \mid \mathcal{A}_0) \leq \max_{[q,d] \in \mathcal{A}^*(w' \mid [(1,0,0),0])} E_q[y] - d$. Since all actions in $\mathcal{A}_0$ except $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ have $E_q[y] - d \leq \frac{72}{100}$, $V_P^D(w'; \mathcal{N}(w)) \leq \frac{72}{100}$ for any w' such that $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}] \notin \mathcal{A}^*(w' \mid [(1, 0, 0), 0])$.

The only contracts w' it remains to consider are those with $[(\frac{16}{100},\frac{24}{100},\frac{60}{100}),\frac{3}{10}] \in \mathcal{A}^*(w' \mid [(1,0,0),0])$ and with $V_P(w' \mid \mathcal{A}_0) > \frac{72}{100}$. Solving for the smallest value of $w'(2)$ such that $[(\frac{16}{100},\frac{24}{100},\frac{60}{100}),\frac{3}{10}] \in \mathcal{A}^*(w' \mid [(1,0,0),0])$ yields that $w'(2) \geq \frac{4737}{6100}$. Solving for the upper bound on $w'(2)$ such that $V_P(w' \mid \mathcal{A}_0) > \frac{72}{100}$ yields that $w'(2) < \frac{6-2w'(1)}{5}$. Consider the agent $[(\frac{47}{100},\frac{1}{100},\frac{52}{100}),0]$. Observe that if the agent took this action in response to w', the principal's profit would be $\frac{52}{100}(2-w'(2))+\frac{1}{100}(1-w'(1)) \leq \frac{24636}{38125} < \frac{65}{100} < \frac{72}{100}$. The agent will indeed respond to w' with $[(\frac{47}{100},\frac{1}{100},\frac{52}{100}),0]$ if and only if $\frac{1}{100}w'(1)+\frac{52}{100}w'(2) > \frac{24}{100}w'(1)+\frac{60}{100}w'(2)-\frac{3}{10}$, which is equivalent to $w'(2) < \frac{30-23w'(1)}{8}$. Since $w'(2) < \frac{6-2w'(1)}{5} \leq \frac{30-23w'(1)}{8}$ for all $w'(1) \in [0,1]$, $V_P(w' \mid \mathcal{A}_0) > \frac{72}{100}$ implies that the agent will indeed respond to w' with $[(\frac{47}{100},\frac{1}{100},\frac{52}{100}),0]$. Furthermore, $[(\frac{47}{100},\frac{1}{100},\frac{52}{100}),0] \in \mathcal{N}(w)$ since, under non-disclosure and contract w, the agent takes the additional action and gets payoff $\frac{3}{2500}$ whereas under disclosure the principal would offer the zero contract and the agent would get 0. Therefore, for any such contract w', $V_P^D(w';\mathcal{N}(w)) < \frac{65}{100} < \frac{72}{100}$. This completes the proof of (2), showing that there is an equilibrium where the non-linear contract $w = (0,\frac{3}{25},0)$ is offered following non-disclosure.

It remains to show part (3) of the result: that there is no equilibrium in which a linear contract is offered following non-disclosure.

Suppose $w(y) = \beta^N y$ for some $\beta^N \geq 0$ is offered in equilibrium following non-disclosure. Observe that for an agent having only $\mathcal{A}_0$ available, the agent's best response to $\beta^N y$ is $[(1,0,0),0]$ for $\beta^N \in [0,\frac{12}{475})$, $[(\frac{22}{32},\frac{1}{32},\frac{9}{32}),\frac{3}{200}]$ for $\beta^N \in [\frac{12}{475},\frac{12}{125})$, $[(\frac{1}{2},\frac{1}{4},\frac{1}{4}),\frac{3}{100}]$ for $\beta^N \in [\frac{12}{125},\frac{9}{23})$,and $[(\frac{16}{100},\frac{24}{100},\frac{60}{100}),\frac{3}{10}]$ for $\beta^N \geq \frac{9}{23}$. Therefore, $V_P(\beta^N y \mid \mathcal{A}_0) = 0$ for $\beta^N \in [0,\frac{12}{475})$, $V_P(\beta^N y \mid \mathcal{A}_0) \leq \frac{463}{800} < \frac{58}{100}$ for $\beta^N \in [\frac{12}{475},\frac{12}{125})$, $V_P(\beta^N y \mid \mathcal{A}_0) \leq \frac{339}{500} < \frac{68}{100}$ for $\beta^N \in [\frac{12}{125},\frac{9}{23})$, and $V_P(\beta^N y \mid \mathcal{A}_0) \leq \frac{504}{575} < \frac{88}{100}$ for $\beta^N \geq \frac{9}{23}$. The remainder of the argument is divided into four cases: $\beta^N \leq \frac{187}{732}$, $\frac{187}{732} < \beta^N < \frac{9}{23}$, $\frac{9}{23} \leq \beta^N \leq \frac{2}{3}$ and $\beta^N > \frac{2}{3}$.

Case 1: $0 \leq \beta^N \leq \frac{187}{732}$. Since $\frac{187}{732} < \frac{9}{23}$, $V_P^D(\beta^N y) \leq V_P(\beta^N y \mid \mathcal{A}_0) \leq \frac{339}{500} < \frac{72}{100}$. Consider a potential deviation to offer $w' = (0,\frac{3}{25},0)$ following non-disclosure. Observe that for an agent having only $\mathcal{A}_0$ available, the best response to the contract w' is $\mathcal{A}^*(w' \mid \mathcal{A}_0) = \{[(\frac{1}{2},\frac{1}{4},\frac{1}{4}),\frac{3}{100}]\}$. Thus, $V_P(w' \mid \mathcal{A}_0) = E_{(\frac{1}{2},\frac{1}{4},\frac{1}{4})}[y] - \frac{3}{100} = \frac{72}{100}$ and therefore $V_P^D(w';\mathcal{N}(\beta^N y)) \leq \frac{72}{100}$. We claim that for $\beta^N \leq \frac{187}{732}$ this bound is tight, i.e., $V_P^D(w';\mathcal{N}(\beta^N y)) = \frac{72}{100}$, implying that w' is a robustly strictly profitable deviation from $\beta^N y$.

Showing this bound is tight is equivalent to showing that for all agents $[p,c]$ such that $E_p[y] - E_p[w'(y)] < \frac{72}{100}$ and $E_p[w'(y)] - c > V_A(w' \mid \mathcal{A}_0) = 0$ (i.e., such that the agent would respond to w' with action $[p,c]$ and this would yield the principal a payoff strictly below $\frac{72}{100}$), $[p,c] \notin \mathcal{N}$. To do this, first, suppose such a $[p,c]$ is such that the principal would implement the action $[(\frac{16}{100},\frac{24}{100},\frac{60}{100}),\frac{3}{10}]$ following disclosure. As was established previously, the principal can get at most $\frac{5679}{6100}$ by implementing $[(\frac{16}{100},\frac{24}{100},\frac{60}{100}),\frac{3}{10}]$ following disclosure of $[p,c]$, and the corresponding payoff to the agent is at least $\frac{51}{244}$. Thus, under our supposition, $V_A(w^{[p,c]} \mid [p,c]) \geq \frac{51}{244} = \frac{187}{732}(\frac{72}{100}+\frac{3}{25}\frac{9}{11}) > \frac{187}{732}E_p[y]-c \geq \beta^N E_p[y]-c$, where the third inequality follows from $E_p[y]-E_p[w'(y)] < \frac{72}{100}$ and the fourth inequality follows from $\beta^N \leq \frac{187}{732}$, and therefore $[p,c] \notin \mathcal{N}$.

Finally, recall from the earlier arguments in the corresponding paragraphs in the proof of part (2) of this theorem that under the assumptions $E_p[y]-E_p[w'(y)] < \frac{72}{100}$ and $E_p[w'(y)] - c > V_A(w' \mid \mathcal{A}_0) = 0$, the principal does strictly better following disclosure of $[p,c]$ by implementing $[(\frac{16}{100},\frac{24}{100},\frac{60}{100}),\frac{3}{10}]$ instead of any of the agent's other actions. This completes the argument that w' is a robustly strictly profitable deviation in Case 1.

Case 2: $\frac{187}{732} < \beta^N < \frac{9}{23}$. First consider any β^N for which allowing disclosure does not increase the principal's worst-case payoff following non-disclosure (i.e., $V_P^D(\beta^N y) = V_P(\beta^N y)$). Then $\frac{1}{5}y$ is a robustly strictly profitable deviation from $\beta^N y$ for the principal, since $V_P^D(\beta^N y) = V_P(\beta^N y) < V_P(\frac{1}{5}y) \leq V_P^D(\frac{1}{5}y)$, where the middle inequality follows from the application in the first paragraph of this proof of Carroll's Lemma 2 to our example.

It remains to consider β^N for which $V_P^D(\beta^N y) > V_P(\beta^N y)$. We show that $\frac{1}{5}y$ is a robustly strictly profitable deviation for the principal following non-disclosure. For $\beta \in \{\frac{1}{5},\beta^N\}$, since $\mathcal{A}^*(\beta y \mid \mathcal{A}_0) = \{[(\frac{1}{2},\frac{1}{4},\frac{1}{4}),\frac{3}{100}]\}$,

$$V_P^D(\beta y;\mathcal{N}(\beta^N y)) = (1-\beta)\inf\{k \mid \exists[q,d]\in\mathcal{N}(\beta^N y) \text{ s.t. } E_q[y]=k \text{ and } k\beta-d \geq \frac{3}{4}\beta-\frac{3}{100}, \text{ and, if equality holds, } k\geq\frac{3}{4}\}. \tag{3.2}$$

Since $\mathcal{A}_0 \subseteq \mathcal{N}(\beta^N y)$, the infimum on the right-hand side of (3.2) is bounded above by $\frac{3}{4}$. Therefore, the right-hand side of (3.2) can be replaced by

$$(1-\beta)\inf\{k \mid \exists[q,d]\in\mathcal{N}(\beta^N y) \text{ s.t. } E_q[y]=k \text{ and } \frac{3}{4}\geq k > \frac{d}{\beta}+\frac{3}{4}-\frac{3}{100\beta} \text{ or } k=\frac{d}{\beta}+\frac{3}{4}-\frac{3}{100\beta}\geq\frac{3}{4}\}. \tag{3.3}$$

By Lemma 3.1, this becomes

$$V_P^D(\beta y;\mathcal{N}(\beta^N y)) = (1-\beta)\inf\{\{k \mid \exists[q,0]\in\mathcal{N}(\beta^N y) \text{ s.t. } E_q[y]=k \text{ and } k > \frac{3}{4}-\frac{3}{100\beta}\}\cup\{\frac{3}{4}\}\}. \tag{3.4}$$

Since, by assumption, $V_P^D(\beta^N y) > V_P(\beta^N y) = (1-\beta^N)(\frac{3}{4} - \frac{3}{100\beta^N})$, then

$$\nexists [q,0] \in \mathcal{N}(\beta^N y) \text{ s.t. } E_q[y] = k \text{ and } k \in (\frac{3}{4} - \frac{3}{100\beta^N}, \frac{V_P^D(\beta^N y)}{1-\beta^N}).$$

It then follows from Lemma 3.2 that there is no $[q,0] \in \mathcal{N}(\beta^N y) \setminus \mathcal{A}_0$ having any lower expected output than $\frac{V_P^D(\beta^N y)}{1-\beta^N}$:

$$\nexists [q,0] \in \mathcal{N}(\beta^N y) \setminus \mathcal{A}_0 \text{ s.t. } E_q[y] = k \text{ and } k \in [0, \frac{V_P^D(\beta^N y)}{1-\beta^N}).$$

Thus, $V_P^D(\frac{1}{5}y; \mathcal{N}(\beta^N y)) = (1-\frac{1}{5}) \min\{\frac{3}{4}, \frac{V_P^D(\beta^N y)}{1-\beta^N}\}$. Since $\frac{V_P^D(\beta^N y)}{1-\beta^N} \leq \frac{V_P(\beta^N y|\mathcal{A}_0)}{1-\beta^N} = \frac{3}{4}$,

$$V_P^D(\frac{1}{5}y; \mathcal{N}(\beta^N y)) = (1-\frac{1}{5})\frac{V_P^D(\beta^N y)}{1-\beta^N} > V_P^D(\beta^N y)$$

and the contract $\frac{1}{5}y$ is therefore a strictly profitable deviation from $\beta^N y$ for the principal following non-disclosure. This completes the proof of Case 2.

Case 3: $\frac{9}{23} \leq \beta^N \leq \frac{2}{3}$. Fix any such β^N, and consider the agent $[p,0] \equiv [(\frac{3}{20\beta^N} - \frac{3}{100}, \frac{60}{100}, \frac{43}{100} - \frac{3}{20\beta^N}), 0]$. Observe that $\mathcal{A}^*(\beta^N y \mid \mathcal{A}_0) = \{[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]\}$ and $\mathcal{A}^*(\beta^N y \mid [p,0]) = \{[p,0]\}$, where the latter equality follows from $\beta^N E_p[y] = \frac{73}{50}\beta^N - \frac{3}{10} > \frac{72}{50}\beta^N - \frac{3}{10} = V_A(\beta^N y \mid \mathcal{A}_0)$. We next show that $[p,0] \in \mathcal{N}(\beta^N y)$. The agent's payoff under non-disclosure is $\beta^N E_p[y] = \frac{73}{50}\beta^N - \frac{3}{10} > 0$. Under disclosure the agent's payoff will depend on what the principal implements and the contract, $w^{[p,0]}$, offered. First, suppose that the principal would implement the action $[q,d] = [(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ following disclosure of $[p,0]$. For $\beta^N \in [\frac{9}{23}, \frac{2}{3}]$, the solution to the program (3.1) with this $[q,d]$ yields a payoff to the principal of $\min\{\frac{5679}{6100}, \frac{69660+257058\beta^N}{62625+300475\beta^N}\}$ and a corresponding payoff to the agent of $\max\{\frac{51}{244}, \frac{3465+170967\beta^N}{125250+600950\beta^N}\}$, which is strictly less than the agent's non-disclosure payoff of $\frac{73}{50}\beta^N - \frac{3}{10}$, implying $[p,0] \in \mathcal{N}(\beta^N y)$.

Second, suppose the principal would implement the action $[p,0]$ following disclosure of $[p,0]$. Then the contract following disclosure, $w^{[p,0]}$, is the zero contract, yielding the agent a payoff of $0 < \frac{73}{50}\beta^N - \frac{3}{10}$, again implying $[p,0] \in \mathcal{N}(\beta^N y)$.

Third, suppose that the principal would implement the action $[q,d] = [(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}]$ following disclosure of $[p,0]$. A solution to the corresponding program (3.1) with this $[q,d]$ exists in Case 3 if and only if $\frac{9}{23} \leq \beta^N \leq \frac{23685}{47297} \approx 0.501$, and yields a payoff to the principal of $\frac{81}{700}\frac{95-159\beta^N}{15-23\beta^N}$, which is less than the principal's payoff of $\min\{\frac{5679}{6100}, \frac{69660+257058\beta^N}{62625+300475\beta^N}\}$ from implementing the action $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ for $\frac{9}{23} \leq \beta^N \leq \frac{23685}{47297}$. Thus the principal does strictly better following disclosure of $[p,0]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ instead of

$[(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}]$.

Fourth, suppose that the principal would implement the action $[q, d] = [(\frac{22}{32}, \frac{1}{32}, \frac{9}{32}), \frac{3}{200}]$ following disclosure of $[p, 0]$. Solving the corresponding program (3.1) with this $[q, d]$ yields a payoff to the principal of $\frac{2280-2369\beta^N}{3840-3808\beta^N}$, which is less than $\min\{\frac{5679}{6100}, \frac{69660+257058\beta^N}{62625+300475\beta^N}\}$, the principal's payoff from implementing the action $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ for $\frac{9}{23} \leq \beta^N \leq \frac{2}{3}$. Thus the principal does strictly better following disclosure of $[p, 0]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ instead of $[(\frac{22}{32}, \frac{1}{32}, \frac{9}{32}), \frac{3}{200}]$.

Finally, since $E_p[y] > 0$, the principal does strictly better following disclosure of $[p, 0]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ instead of $[(1, 0, 0), 0]$.

Thus for Case 3 we have established that $[p, 0] \in \mathcal{N}(\beta^N y)$, implying $V_P^D(\beta^N y) \leq V_P(\beta^N y \mid [p, 0])$. Therefore, $V_P^D(\beta^N y) \leq V_P(\beta^N y \mid [p, 0]) = (1 - \beta^N)(\frac{73}{50} - \frac{3}{10\beta^N}) \leq \max_{\alpha \in [\frac{9}{23}, \frac{2}{3}]}(1 - \alpha)(\frac{73}{50} - \frac{3}{10\alpha}) \leq \frac{44}{100} < \frac{12}{25} = V_P(\frac{1}{5}y) \leq V_P^D(\frac{1}{5}y; \mathcal{N}(\beta^N y))$, and so the contract $\frac{1}{5}y$ is a robustly strictly profitable deviation in Case 3.

Case 4: $\beta^N > \frac{2}{3}$. In this case, $V_P^D(\beta^N y) \leq V_P(\beta^N y \mid \mathcal{A}_0) = (1 - \beta^N)\frac{36}{25} < \frac{12}{25} = V_P(\frac{1}{5}y) \leq V_P^D(\frac{1}{5}y; \mathcal{N}(\beta^N y))$. Therefore the contract $\frac{1}{5}y$ is a robustly strictly profitable deviation in Case 4. □

Lemma 3.1. *Suppose a positive linear contract $\beta^N y$ is offered following non-disclosure, and that, for some $k \in [0, \frac{3}{4}]$, there exists a non-disclosing additional action $[p, c] \notin \mathcal{A}_0$ with $E_p[y] = k$ and $c > 0$. Then, there also exists some non-disclosing additional action $[p', 0]$ with $E_{p'}[y] = k$.*

Proof. First consider the comparison of the payoffs following non-disclosure. Recall that the agent is offered the contract $w(y) = \beta^N y$. Consider any additional action $[p', 0]$ such that $E_{p'}[y] = k$. Since $\beta^N k > \beta^N k - c$ so that the only change in the agent's choice problem following non-disclosure is replacing the action $[p, c]$ by an action $[p', 0]$ having better payoff under the contract $\beta^N y$, an agent with $[p', 0]$ would have expected payoff following non-disclosure at least as high as an agent with $[p, c]$. Now turn to the comparison following disclosure. There are five cases corresponding to the different actions that the principal might want to implement following disclosure of $[p, c]$.

Consider any $k \in [0, \frac{3}{4}]$ such that the antecedent in the statement of the lemma is satisfied. Suppose suppose that the principal would implement the action $[q, d] = [(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ following disclosure of $[p, c]$. As was established in the proof of Theorem 3.1, the principal can get at most $\frac{5679}{6100}$ by implementing this $[q, d]$ following disclosure of $[p, c]$, and the corresponding payoff to the agent is at least $\frac{51}{244}$. Let $p' = (1 - k, k, 0)$. If the agent discloses $[p', 0]$ then, by solving program (3.1) with this $[q, d]$ and additional action $[p', 0]$ and noting the solution, $\frac{5679}{6100}$, is larger than $\max_{[q,d] \in (\mathcal{A}_0 \setminus [(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]) \cup \{[p', 0]\}} E_q[y]$, we establish that the principal best responds to disclosure of $[p', 0]$ by implementing $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$, yielding the agent a payoff of $\frac{51}{244}$ that is weakly less than the agent's payoff from disclosing $[p, c]$, as desired.

Next, suppose instead that $[p, c]$ is such that, if $[p, c]$ were to be disclosed, the principal would best respond by offering a wage contract that implements $[p, c]$. Let $p' = p$. Observe that for all $c' \leq c$, any wage contract that implements $[p, c]$ also implements $[p, c']$, and any wage contract that implements a particular action in $\mathcal{A}_0$ following disclosure of $[p, c']$ implements the same action following disclosure of $[p, c]$. Therefore, the principal's payoff from implementing $[p, c]$, if it were to be disclosed, is non-increasing in c, and the principal's payoff from implementing any action in $\mathcal{A}_0$ is non-decreasing in c. Since the principal implements $[p, c] \notin \mathcal{A}_0$ following disclosure, they will also find it optimal to implement $[p, 0]$, which can be done with the zero contract. Therefore the agent's payoff from disclosing $[p', 0]$ with $p' = p$ is zero, which is weakly less than their payoff from disclosing $[p, c]$, as desired.

Next, suppose instead that $[p, c]$ is such that, if $[p, c]$ were to be disclosed, the principal would best respond by offering a wage contract that implements $[q, d] = [(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}]$. As was established in the proof of Theorem 3.1, the principal can get at most $\frac{72}{100}$ by implementing this $[q, d]$ following disclosure of $[p, c]$. Tightening program (3.1) with $[q, d] = [(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ by replacing the right-hand side of the constraint involving the additional action by the upper bound $kw(1) + \max\{0, \frac{k}{2}(w(2) - 2w(1))\}$, imposing $k \leq \frac{3}{4}$ and solving, yields a lower bound for the principal's payoff as a function of k. Minimizing over $k \in [0, \frac{3}{4}]$ yields a lower bound of $\frac{468}{575}$. Since $\frac{468}{575} > \frac{72}{100}$, the principal prefers to implement $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$ rather than $[(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}), \frac{3}{100}]$ following disclosure.

Next, suppose instead that $[p, c]$ is such that, if $[p, c]$ were to be disclosed, the principal would best respond by offering a wage contract that implements $[q, d] = [(\frac{22}{32}, \frac{1}{32}, \frac{9}{32}), \frac{3}{200}]$. As was established in the proof of Theorem 3.1, the principal can get at most $\frac{463}{800}$ by implementing this $[q, d]$ following disclosure of $[p, c]$. Since $\frac{468}{575} > \frac{463}{800}$, the principal will not want to implement $[(\frac{22}{32}, \frac{1}{32}, \frac{9}{32}), \frac{3}{200}]$ following disclosure.

Finally, suppose that $[p, c]$ is such that, if $[p, c]$ were to be disclosed, the principal would best respond by offering a wage contract that implements $[(1, 0, 0), 0] \in \mathcal{A}_0$. The principal gets 0 by doing so. Since $\frac{468}{575} > 0$, the principal will not want to implement $[(1, 0, 0), 0]$ following disclosure.

This completes the proof, as we have shown that for $k \in [0, \frac{3}{4}]$, $[p', 0]$ for either $p' = p$ or $p' = (1 - k, k, 0)$ would not be disclosed by the agent. □

Lemma 3.2. *Suppose a positive linear contract $\beta^N y$ is offered following non-disclosure, and that, for some $k \in [0, \frac{72}{100})$, there exists a non-disclosing agent $[p, 0] \notin \mathcal{A}_0$ with $E_p[y] = k$. Then, for any $k' \in (k, \frac{72}{100}]$ there exists some non-disclosing agent $[p', 0]$ with expected output k'.*

Proof. Fix any $k' \in (k, \frac{72}{100}]$. We will show that there exists an agent with $[p', 0]$ who, compared to an agent with the additional action $[p, 0]$, has a weakly higher payoff following non-disclosure and a weakly lower payoff following disclosure. This implies that since an agent with $[p, 0] \notin \mathcal{A}_0$ is assumed not to disclose, the same will be true of an agent with such a $[p', 0]$.[5] First consider the comparison of the payoffs following non-disclosure. Recall that the agent is offered the contract $w(y) = \beta^N y$. Consider any additional action $[p', 0]$ such that $E_{p'}[y] = k'$. Since $\beta^N k' > \beta^N k$ so that the only change in the agent's choice problem following non-disclosure is replacing the action $[p, 0]$ by an action $[p', 0]$ having strictly better payoff under the contract $\beta^N y$, an agent with $[p', 0]$ would have expected payoff following non-disclosure at least as high as an agent with $[p, 0]$.

Next turn to the comparison of payoffs following disclosure. Consider any $k \in [0, \frac{72}{100})$ such that the antecedent in the statement of the lemma is satisfied. Suppose the additional action $[p, 0]$ is such that, if this additional action were to be disclosed, the principal would best respond by offering a wage contract that implements the action $[q, d] = [(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$. As was established in the proof of Theorem 3.1, the principal can get at most $\frac{5679}{6100}$ by implementing this $[q, d]$ following disclosure of $[p, 0]$, and the corresponding payoff to the agent is at least $\frac{51}{244}$. Suppose $k' \in (k, \frac{72}{100}]$ and consider $p' = (1 - k', k', 0)$. This action has expected output k'. If the agent were to disclose $[p', 0]$, and the principal wants to implement $[(\frac{16}{100}, \frac{24}{100}, \frac{60}{100}), \frac{3}{10}]$,then, as was shown in the proof of Lemma 3.1, this yields a payoff to the principal of $\frac{5679}{6100}$ and to the agent of $\frac{51}{244}$. If, instead, the principal wants to implement $[p', 0]$,they do so using the zero contract, giving the agent a payoff of zero and the principal k'. The arguments in the last few paragraphs of the proof of Lemma 3.1 demonstrate that the principal will not want to implement any of the remaining actions in $\mathcal{A}_0$. Since, compared to the $[p, 0]$ case, we have now shown that for additional action $[p', 0]$ with $p' = (1 - k', k', 0)$ and $k' \in (k, \frac{72}{100}]$, the agent's non-disclosure payoff is weakly higher and disclosure payoff is weakly lower, the fact that $[p, 0]$ did not disclose implies that this $[p', 0]$ will not disclose.

Next, suppose that if $[p, 0]$ were to be disclosed, the principal would best respond by offering a wage contract that implements $[p, 0]$. The best such contract for the principal is the zero contract, and the payoff to the principal is k, while the agent gets zero. Consider some additional action $[p', 0]$ such that $E_{p'}[y] = k'$, $p'(1) \geq p(1)$ and $p'(2) \geq p(2)$. The agent's expected payoff from $[p', 0]$ is always at least as high as from $[p, 0]$ for all wage contracts (since $w(0) = 0$). Thus, the principal's payoff following disclosure from implementing any action in $\mathcal{A}_0$ cannot increase compared to what it was with $[p, 0]$ since the incentive constraints changed by the additional action can only become tighter. Additionally, the principal's

[5] If $[p, 0] \in \mathcal{A}_0$, this logic is no longer valid, as additional actions in $\mathcal{A}_0$ cannot be disclosed even if the agent would wish to do so.

payoff following disclosure from implementing the additional action using the zero contract increases to k' from k. Thus, the assumption that the principal implements $[p, 0]$ following disclosure implies that the principal also wants to implement $[p', 0]$ following disclosure. Therefore, as desired, the agent's payoff following disclosure of $[p', 0]$ does not increase over that for $[p, 0]$, and therefore $[p', 0]$ will not be disclosed.

The arguments in the last few paragraphs of the proof of Lemma 3.1 demonstrate that the principal will not want to implement any of the remaining actions in $\mathcal{A}_0$ if $[p, 0]$ were to be disclosed. This completes the proof.

□

3.1.1 What if an agent with only $\mathcal{A}_0$ available can choose to verifiably disclose that?

In this section we consider the possibility that an agent having only actions in $\mathcal{A}_0$ available (i.e., $[p, c] \in \mathcal{A}_0$) is able, contrary to the assumption of our main model, to verifiably disclose that no other actions are available. Consequently we must also modify the definition of an equilibrium (Definition 2.6) by requiring conditions (ii) and (iii) of that definition to apply for all $[p, c]$, not just $[p, c] \notin \mathcal{A}_0$. The next result shows that this change in assumptions leads to a very different conclusion about the existence of equilibria with a linear contract offered following non-disclosure: there is always an equilibrium where some non-disclosure occurs and the zero contract is offered following non-disclosure.

Theorem 3.2. *For all commonly known technologies $\mathcal{A}_0$, when voluntary disclosure, including of $\mathcal{A}_0$, is allowed, there is an equilibrium with a non-empty non-disclosure set in which the robust contract offered following non-disclosure is the zero contract.*

Proof. Suppose the principal offers the zero contract following non-disclosure, i.e., $w^{\emptyset} = 0^{\mathcal{Y}}$. Let $\mathcal{N}(0^{\mathcal{Y}}) = \{[\delta_{\max \mathcal{Y}}, 0]\}$ be the non-disclosure set. Since $V_A(w^{[\delta_{\max \mathcal{Y}}, 0]} \mid \mathcal{A}_0 \bigcup \{[\delta_{\max \mathcal{Y}}, 0]\}) = 0 = V_A(0^{\mathcal{Y}} \mid \mathcal{A}_0 \bigcup \{[\delta_{\max \mathcal{Y}}, 0]\})$, non-disclosure is an optimal choice for $[\delta_{\max \mathcal{Y}}, 0]$. Moreover, since $V_A(w^{[p,c]} \mid \mathcal{A}_0 \bigcup \{[p, c]\}) \geq E_q[w^{[p,c]}(y)] - d \geq -d = V_A(0^{\mathcal{Y}} \mid \mathcal{A}_0 \bigcup \{[p, c]\})$ for any $[q, d] \in \mathcal{A}^*(0^{\mathcal{Y}} \mid [p, c])$, disclosure is an optimal choice for all $[p, c] \neq [\delta_{\max \mathcal{Y}}, 0]$. Therefore the set $\mathcal{N}(0^{\mathcal{Y}})$ satisfies condition (ii) of Definition 2.6 for all $[p, c]$. Furthermore, $0^{\mathcal{Y}} \in \arg\max_{w \in \mathbb{W}} V_P(w \mid [\delta_{\max \mathcal{Y}}, 0]) = \arg\max_{w \in \mathbb{W}} V_P^D(w; \mathcal{N})$, implying that condition (iii) of Definition 2.6 is satisfied. This completes the proof. □

Several remarks about Theorem 3.2 are worth noting. First, the existence of such an equilibrium (i.e., an equilibrium in which the zero contract is offered following non-disclosure, and only the agent $[\delta_{\max \mathcal{Y}}, 0]$ does not disclose) is quite robust to the assumptions about how many additional actions an agent might have available to disclose. In particular, our assumption that no agent has more than a single additional action beyond $\mathcal{A}_0$ is inessential for Theorem 3.2 – in the proof, simply replace $w^{[p,c]}$ with $w^{\mathcal{B}}$, the contract the principal offers in response to the disclosure of a compact set of actions $\mathcal{B} \subseteq \mathcal{A}$ and observe that no matter what such contracts are, some (full or partial) disclosure remains an optimal choice for all agents such that their available set of actions, $\mathcal{A}$, is something other than $\mathcal{A}_0 \bigcup \{[\delta_{\max \mathcal{Y}}, 0]\}$.

Second, the fully-revealing equilibrium described in Theorem 3.2 results in the same contracts and outcomes as in an equilibrium of the game where the agent's available actions were fully observable from the start. In this sense, Theorem 3.2 says that the possibility of verifiable disclosure, when including the possibility of disclosing that no additional actions are available, can completely undo any effects of the uncertainty the principal faces about the agent's available actions. For the same reason, this is the principal's most-preferred equilibrium of the game. Furthermore, the statements in this paragraph remain true even if agents may have more than one additional action as long as the agent's disclosure decision is limited to fully disclosing or not disclosing (i.e., partial disclosure is not permitted).

3.2 Sufficient Conditions for Non-linearity

In this section, we provide sufficient conditions under which a non-linear contract is offered in equilibrium following non-disclosure. We begin by listing some assumptions on $\mathcal{A}_0$ (in addition to the assumption used throughout that it contains at least one action with $E_q[y] - d > 0$) that will be our starting point in trying to generalize the example used in proving Theorem 3.1. The first two assumptions are meant for simplicity (e.g., the finiteness of $\mathcal{A}_0$) and to describe conditions, such as the existence of a zero-cost, zero-output action and convexity of costs for increasing expected output, that are common in the Principal-Agent context. The third and fourth assumptions describe what we think are key features of the example used in proving Theorem 3.1, especially relating to the distinction between what would be implemented under the robustly optimal linear contract when disclosure is not allowed and what would be implemented if it were common knowledge that only the actions in $\mathcal{A}_0$ were available to the agent. The fifth and final assumption contains the remaining conditions that we use in the current sufficiency argument. We remark that while the conditions in the fifth assumption are not difficult to state, we do not yet have an insightful interpretation of them.

Assumption 1. $\mathcal{A}_0$ *consists of a finite number of actions and includes the zero action,* $(\delta_0, 0)$ *and* $|\mathcal{Y}| \geq 3$.

Assumption 2. *There exists a function* $C : \{E_q[y] \mid [q,d] \in \mathcal{A}_0\} \to \mathbb{R}_+$ *that is strictly increasing, strictly convex, and for which* $y - C(y)$ *is strictly increasing, such that* $[q,d] \in \mathcal{A}_0$ *implies* $d = C(E_q[y])$*, i.e., across* $\mathcal{A}_0$*, costs and surplus are strictly increasing in expected output and additional expected output is increasingly costly.*

Assumption 3. *If the principal knew that only actions in* $\mathcal{A}_0$ *were available to the agent, it is optimal for the principal to offer a contract* w^0 *that implements a highest surplus action in* $\mathcal{A}_0$*.*

Denote by $[q^0, d^0]$ the action taken by an agent with only actions in $\mathcal{A}_0$ available if offered contract w^0 by the principal.

Assumption 4. *All positive expected output actions in* $\mathcal{A}_0$ *have full support on* $\mathcal{Y}$*, the set* $\arg\max_{[r,e]\in\mathcal{A}_0} \sqrt{E_r[y]} - \sqrt{e}$ *is a singleton set consisting of an action with positive cost, the linear contract* βy *with* $\beta = \sqrt{\frac{d}{E_q[y]}}$ *for* $[q,d] \in \arg\max_{[r,e]\in\mathcal{A}_0} \sqrt{E_r[y]} - \sqrt{e}$ *does not implement a highest surplus action in* $\mathcal{A}_0$*, there is a least-cost contract implementing the same action in* $\mathcal{A}_0$ *as* βy *that is non-linear, and an agent having only* $\mathcal{A}_0$ *available is strictly better off under contract* w^0 *than under contract* βy*.*

Assumption 5. *There is a non-linear contract* w *that is a least-cost way to implement the same action in* $\mathcal{A}_0$ *as* βy *given that only* $\mathcal{A}_0$ *is available such that* $V_P^D(w) > 0$*. Additionally, for any contract* w' *such that* $V_P(w' \mid \mathcal{A}_0) > V_P^D(w)$*, there exist* $r \in \Delta(\mathcal{Y})$ *and* $\varepsilon > 0$ *such that* $E_r[y - w'(y)] > 0$*,* $\varepsilon \le \min_{y\in\mathcal{Y}} \frac{r(y)V_P^D(w)}{E_r[y-w'(y)]}$*, and the following hold:*

$$(i) E_r[y - w'(y)] \ge V_P^D(w),$$
$$(ii) E_{q^0}[y - w^0(y)] + \varepsilon \le \frac{E_r[y]}{E_r[y - w'(y)]} V_P^D(w), \text{ and}$$
$$(iii) \max_{[q,d]\in\mathcal{A}_0} E_q[w'(y)] - d + \varepsilon \le \frac{E_r[w'(y)]}{E_r[y - w'(y)]} V_P^D(w).$$

Theorem 3.3. *Under Assumptions 1-5, (1) when no disclosure is allowed, the unique equilibrium robust contract is a positive linear contract, and (2) when voluntary disclosure is allowed, there is an equilibrium in which the robust contract offered following non-disclosure is a non-linear contract.*

Proof of Theorem 3.3. Suppose first that no disclosure is allowed. By Carroll (2015), since under Assumption 4 all positive expected output actions in $\mathcal{A}_0$ have full support on $\mathcal{Y}$, any equilibrium robust contract is linear. Furthermore, such linear contracts have coefficient $\beta = \sqrt{\frac{d}{E_q[y]}}$ for $[q,d] \in \arg\max_{[r,e]\in\mathcal{A}_0} \sqrt{E_r[y]} - \sqrt{e}$ and $V_P(\beta y) > 0$. Assumption 4 implies there is a unique such β and that it is positive. This establishes (1).

Now consider the game where disclosure is allowed. Since disclosure can only remove some additional actions from the principal's consideration following non-disclosure, allowing disclosure can never strictly lower the worst-case payoff of a contract following non-disclosure, meaning $V_P(w) \leq V_P^D(w)$ for any contract w. Furthermore, since an agent with only the actions in $\mathcal{A}_0$ available can never disclose, an upper bound on $V_P^D(w)$ is given by the principal's payoff from offering w to an agent having only $\mathcal{A}_0$ available. Thus $V_P^D(w) \leq V_P(w \mid \mathcal{A}_0)$. Note that if $V_P^D(w) = 0$ then it cannot be part of an equilibrium for the principal to offer w following non-disclosure. This follows since $V_P^D(\beta y) \geq V_P(\beta y) > 0$ would then imply that offering βy would be a robustly strictly profitable deviation from w following non-disclosure.

Consider the non-linear contract w referred to in Assumption 5. By that assumption, $V_P^D(w) > 0$. We next show that the principal offering w following non-disclosure is part of an equilibrium by showing that there is no other contract w' that is a robustly strictly profitable deviation from w following non-disclosure. Formally, we want to show that for each $w' \neq w$, $V_P^D(w';\mathcal{N}(w)) \leq V_P^D(w)$. For any contract w', $V_P^D(w';\mathcal{N}(w)) \leq V_P(w' \mid \mathcal{A}_0)$. Thus for any contract w' such that $V_P(w' \mid \mathcal{A}_0) \leq V_P^D(w)$, $V_P^D(w';\mathcal{N}(w)) \leq V_P^D(w)$ follows immediately. To complete the argument therefore, it remains to consider only contracts w' with $V_P(w' \mid \mathcal{A}_0) > V_P^D(w)$.

By Assumption 2, there is a unique maximal surplus level within $\mathcal{A}_0$ and all actions in $\mathcal{A}_0$ attaining that maximum have the same maximal expected output and cost, but may differ in their output distributions. Assumption 3 says that if the principal knew that only actions in $\mathcal{A}_0$ were available to the agent, the principal would offer a contract, denoted w^0, implementing an action, which we denote $[q^0, d^0]$, from among such maximizers. To show that any contract w' with $V_P(w' \mid \mathcal{A}_0) > V_P^D(w)$ has $V_P^D(w';\mathcal{N}(w)) \leq V_P^D(w)$, we will exhibit, for each such w', an additional action $[p, 0]$ such that $[p, 0] \in \mathcal{N}(w) \cap \mathcal{A}^*(w' \mid [p, 0])$ and $E_p[y - w'(y)] \leq V_P^D(w)$. Our strategy for proving such a p exists for each such w' is as follows: First, to guarantee that $[p, 0] \in \mathcal{N}(w)$ we limit attention to p with full support on positive outputs so that the agent's payoff from action $[p, 0]$ under w is positive (thus ensuring that the agent's payoff following non-disclosure is positive when $[p, 0]$ is available), and such that $E_p[y]$ is high enough so that following disclosure the principal would best respond by offering the zero contract, yielding the agent a zero payoff. Second, impose a strict inequality implying $[p, 0] \in \mathcal{A}^*(w' \mid [p, 0])$ and directly require $E_p[y - w'(y)] \leq V_P^D(w)$. Finally, we write a LP to identify such a p with minimal expected output, and show existence by showing that a relaxation of its dual program has a bounded value.

Formally, given w' satisfying $V_P(w' \mid \mathcal{A}_0) > V_P^D(w)$, we will show that there exists a solution to the following LP, where $\varepsilon > 0$ is a parameter used to enforce strictness of selected constraints:

$$\begin{aligned}
&\min_{p\in\mathbb{R}^{\mathcal{Y}}} E_p[y] \\
&\quad \text{s.t.} \\
&\qquad E_p[y] \geq E_{q^0}[y - w^0(y)] + \varepsilon \\
&\qquad - E_p[y - w'(y)] \geq -V_P^D(w) \\
&\qquad E_p[w'(y)] \geq \max_{[q,d]\in\mathcal{A}_0} E_q[w'(y)] - d + \varepsilon \\
&\qquad E_p[1] = 1 \\
&\qquad p \geq \varepsilon.
\end{aligned} \tag{3.5}$$

The first constraint implies that the principal would best respond following disclosure of $[p, 0]$ by offering the zero contract. To see this, first note that by offering the zero contract the principal will get the highest (because of the tie-breaking rule) expected output among all of the agent's available zero cost actions, and that this is bounded below by $E_p[y]$ since $[p, 0]$ is available. Since the zero contract pays no wages, it is the best way for the principal to implement a zero cost action. Furthermore, the principal's payoff from implementing any action in $\mathcal{A}_0$ when additional action $[p, 0]$ is present is bounded above by the principal's payoff from implementing that action ignoring the possibility that the agent has action $[p, 0]$ available, and, by definition of q^0 and w^0, the latter payoff is bounded above by $E_{q^0}[y-w^0(y)]$. The second constraint directly says that the principal's profit under p and contract w' is at most the principal's worst-case payoff following non-disclosure under contract w. The third constraint ensures that an agent with additional action $[p, 0]$ will choose that action if faced with contract w'. The final two constraints ensure that p is a well-defined, full-support output distribution. The choice of minimizing $E_p[y]$ as the objective is purely for convenience, in that it yields a dual that proves tractable to analyze.

The corresponding dual program, where the variables are the multipliers on the primal constraints, specifically the λ are for the output/wage constraints, μ for the equality constraint, and η for the full support constraints, is:

$$\begin{aligned}
&\max_{\lambda\in\mathbb{R}^3,\mu\in\mathbb{R},\eta\in\mathbb{R}^{\mathcal{Y}}} \lambda_1(E_{q^0}[y - w^0(y)] + \varepsilon) - \lambda_2 V_P^D(w) + \lambda_3(\max_{[q,d]\in\mathcal{A}_0} E_q[w'(y)] - d + \varepsilon) + \mu + \varepsilon \sum_{y\in\mathcal{Y}} \eta_y \\
&\quad \text{s.t.} \\
&\qquad \lambda_1 y - \lambda_2(y - w'(y)) + \lambda_3 w'(y) + \mu + \eta_y \leq y, \text{ for all } y \in \mathcal{Y} \\
&\qquad \lambda \geq 0, \eta \geq 0.
\end{aligned} \tag{3.6}$$

Recall that the primal constraints have no solution if and only if the value of the dual is unbounded. We analyze the dual to get insight into the existence of a feasible p for the primal. In particular, we will analyze a relaxed version of the dual in order to provide sufficient conditions for existence of a p satisfying the primal constraints, which, in turn, is sufficient to show that w' is not a strictly profitable deviation from w for the principal.

First, observe that, since $w'(0) = 0$, the dual constraint for $y = 0$ simplifies to $\mu + \eta_0 \leq 0$. Second, consider the relaxation of the dual replacing the constraints for each y with a single constraint formed by taking the expectation of the constraints for the individual y with respect to some $r \in \Delta(\mathcal{Y})$:

$$\max_{\lambda \in \mathbb{R}^3, \mu \in \mathbb{R}, \eta \in \mathbb{R}^{\mathcal{Y}}} \lambda_1(E_{q^0}[y - w^0(y)] + \varepsilon) - \lambda_2 V_P^D(w) + \lambda_3(\max_{[q,d] \in \mathcal{A}_0} E_q[w'(y)] - d + \varepsilon) + \mu + \varepsilon \sum_{y \in \mathcal{Y}} \eta_y$$

s.t.

$$\lambda_1 E_r[y] - \lambda_2 E_r[y - w'(y)] + \lambda_3 E_r[w'(y)] + \mu + E_r[\eta_y] \leq E_r[y] \tag{3.7}$$

$$\lambda \geq 0, \eta \geq 0, \mu + \eta_0 \leq 0.$$

Lemma 3.3 completes the proof by showing that the conditions in Assumption 5 imply the relaxed dual has a bounded solution, and therefore the primal program has a solution, and thus no such w' is a robustly strictly profitable deviation from w. Therefore there is an equilibrium where the principal offers the non-linear contract w following non-disclosure. □

Lemma 3.3. *Suppose $V_P(w' \mid \mathcal{A}_0) > V_P^D(w) > 0$, $E_r[y - w'(y)] > 0$ and $0 < \varepsilon \leq \min_y \frac{r(y) V_P^D(w)}{E_r[y - w'(y)]}$. Program (3.7) has a bounded solution if and only if :*

$$(i) E_r[y - w'(y)] \geq V_P^D(w),$$

$$(ii) E_{q^0}[y - w^0(y)] + \varepsilon \leq \frac{E_r[y]}{E_r[y - w'(y)]} V_P^D(w), \text{ and}$$

$$(iii) \max_{[q,d] \in \mathcal{A}_0} E_q[w'(y)] - d + \varepsilon \leq \frac{E_r[w'(y)]}{E_r[y - w'(y)]} V_P^D(w).$$

Proof. The expectational constraint in (3.7) can be re-written as

$$\lambda_2 \geq \frac{\lambda_1 E_r[y] + \lambda_3 E_r[w'(y)] + \mu + E_r[\eta_y] - E_r[y]}{E_r[y - w'(y)]}.$$

Combining with the non-negativity constraint on λ_2 yields

$$\lambda_2 \geq \max\{0, \frac{\lambda_1 E_r[y] + \lambda_3 E_r[w'(y)] + \mu + E_r[\eta_y] - E_r[y]}{E_r[y - w'(y)]}\}$$

as the only constraint on λ_2. Since the coefficient on λ_2 in the objective function is negative (as $V_P^D(w) > 0$), any bounded solution must have $\lambda_2 = \max\{0, \frac{\lambda_1 E_r[y]+\lambda_3 E_r[w'(y)]+\mu+E_r[\eta_y]-E_r[y]}{E_r[y-w'(y)]}\}$. Substitute that into the objective function for λ_2 allowing us to remove the expectational constraint and the non-negativity constraint on λ_2. Observe that at any bounded solution, $\max\{0, \frac{\lambda_1 E_r[y]+\lambda_3 E_r[w'(y)]+\mu+E_r[\eta_y]-E_r[y]}{E_r[y-w'(y)]}\} = \frac{\lambda_1 E_r[y]+\lambda_3 E_r[w'(y)]+\mu+E_r[\eta_y]-E_r[y]}{E_r[y-w'(y)]}$ since, if not, the objective can be increased by raising λ_1 to the point where $\frac{\lambda_1 E_r[y]+\lambda_3 E_r[w'(y)]+\mu+E_r[\eta_y]-E_r[y]}{E_r[y-w'(y)]} = 0$. The total coefficient on μ in the substituted-into objective function is now $1 - \frac{V_P^D(w)}{E_r[y-w'(y)]}$. Suppose that (i) is violated, and thus that $1 - \frac{V_P^D(w)}{E_r[y-w'(y)]} < 0$. Then, we can unboundedly increase the objective function by simultaneously raising λ_1 while setting $\mu = E_r[y] - \lambda_1 E_r[y] - \lambda_3 E_r[w'(y)] - E_r[\eta_y] \leq -\eta_0$. Thus (i) is necessary for existence of a bounded solution to (3.7) and is assumed for the rest of the argument. By (i), the total coefficient on μ, $1 - \frac{V_P^D(w)}{E_r[y-w'(y)]}$, is non-negative. Therefore, any bounded solution must have $\mu = -\eta_0$ since the only constraint on μ is $\mu \leq -\eta_0$. Substituting, the existence of a bounded solution to (3.7) is equivalent to the existence of a bounded solution to the following:

$$
\begin{aligned}
\max_{\lambda_1,\lambda_3\in\mathbb{R}} & \lambda_1(E_{q^0}[y - w^0(y)] + \varepsilon - \frac{E_r[y]}{E_r[y - w'(y)]}V_P^D(w)) + \\
& \lambda_3((\max_{[q,d]\in\mathcal{A}_0} E_q[w'(y)] - d + \varepsilon) - \frac{E_r[w'(y)]}{E_r[y - w'(y)]}V_P^D(w)) \\
& - \eta_0(1 - \frac{V_P^D(w)}{E_r[y - w'(y)]}) - \frac{E_r[\eta_y]}{E_r[y - w'(y)]}V_P^D(w) + \varepsilon\sum_{y\in\mathcal{Y}}\eta_y \\
\text{s.t.} & \\
& \lambda_1 E_r[y] + \lambda_3 E_r[w'(y)] - \eta_0 + E_r[\eta_y] \geq E_r[y] \\
& \lambda_1, \lambda_3, \eta \geq 0.
\end{aligned}
$$

Since $0 < \varepsilon \leq \min_y \frac{r(y)V_P^D(w)}{E_r[y-w'(y)]}$, the total coefficient on each η_y in the objective function is non-positive and the total coefficient of η_0 in the constraint is negative while the total coefficient of the other η_y in the constraint are positive. Therefore $\eta_0 = 0$ in any bounded solution. Observe that (ii) and (iii) are necessary and sufficient for the coefficients on λ_1 and λ_3 in this objective function to be non-positive. Note that r is strictly positive since $0 < \varepsilon \leq \min_y \frac{r(y)V_P^D(w)}{E_r[y-w'(y)]}$, implying positivity of the coefficients of λ_1 and η_y for $y > 0$ in the constraint; furthermore, $E_r[w'(y)] > 0$ since w' cannot be the zero contract (if $w' = 0$, then $V_P(w' \mid \mathcal{A}_0) > V_P^D(w) > 0$ is only possible if there is a positive expected output action with zero cost in $\mathcal{A}_0$, but this would contradict Assumptions 1 and 2). Given the positivity of the coefficients of λ_1, λ_3 and η_y for $y > 0$ in the constraint, the non-positivity of the coefficients of these variables in the objective function is equivalent to the existence of a bounded solution. □

Proposition 3.1. *The example in the proof of Theorem 3.1 satisfies Assumptions 1-5.*

Proof. TBA □

3.3 Linear robustly optimal contracts even under voluntary disclosure

Having shown that voluntary disclosure can lead all robustly optimal contracts to be non-linear, it is natural to ask if this always occurs. Theorem 3.4 shows that the answer is no, and provides sufficient conditions for there to be an equilibrium in which a linear contract is offered following non-disclosure. In particular, linearity results whenever there are few publicly known-to-be-available actions that generate a positive surplus.

Theorem 3.4. *Suppose among the actions in the commonly known technology $\mathcal{A}_0$ only one, denoted by $[s, f]$, has a positive surplus. Further suppose $f > 0$, and that all positive cost actions in $\mathcal{A}_0$ have full support on $\mathcal{Y}$. Then, (1) when no disclosure is allowed, the unique equilibrium robust contract is a positive linear contract, and (2) when voluntary disclosure is allowed, there is an equilibrium in which the robust contract offered following non-disclosure is the same positive linear contract.*

Proof. When no disclosure is allowed, by Carroll (2015), if all positive cost actions in $\mathcal{A}_0$ have full support on $\mathcal{Y}$, then any equilibrium robust contract is linear. Furthermore, such linear contracts have coefficient $\beta = \sqrt{\frac{d}{E_q[y]}}$ for $[q, d] \in \arg\max_{[r,e]\in\mathcal{A}_0} \sqrt{E_r[y]} - \sqrt{e}$. Observe that the argmax in the previous sentence is the assumed positive surplus action in $\mathcal{A}_0$, implying that this action determines the unique such $\beta = \sqrt{\frac{f}{E_s[y]}} > 0$. This establishes (1).

Now consider the game where disclosure is allowed. Since disclosure can only remove some additional actions from the principal's consideration following non-disclosure, allowing disclosure can never strictly lower the worst-case payoff of a contract following non-disclosure, meaning $V_P(w) \le V_P^D(w)$ for any contract w.

Suppose $w(y) = \beta y$ is offered in equilibrium following non-disclosure. From Carroll (2015), $V_P(\beta y) = \frac{1-\beta}{\beta} V_A(\beta y \mid \mathcal{A}_0)$, which equals $\frac{1-\beta}{\beta}(\beta E_s[y] - f)$ under the assumptions of the theorem. We will now show that $V_P^D(\beta y) = V_P(\beta y)$. Consider an agent $[p'(\varepsilon), 0]$ for $p'(\varepsilon) = (1+\varepsilon)(1-\beta)s + (\beta - \varepsilon(1-\beta))\delta_0$, which is a probability distribution for $\varepsilon \in [0, \frac{\beta}{1-\beta}]$. Observe that, since $\beta = \sqrt{\frac{f}{E_s[y]}}$, for $\varepsilon \in (0, \frac{\beta}{1-\beta})$, $\beta E_{p'(\varepsilon)}[y] > \beta E_{p'(0)}[y] = \beta E_s[y] - f$, where the left-hand side is the agent's payoff following non-disclosure under the additional action and the final term is the corresponding payoff from action $[s, f]$. Thus, if such an agent does not disclose, they take their additional action and their payoff is $\beta E_{p'(\varepsilon)}[y] = \beta(1+\varepsilon)(1-\beta)E_s[y]$. If $\varepsilon \in (0, \frac{\beta}{1-\beta})$ and the agent discloses, we next show that the principal offers the zero contract, implementing the additional action and yielding the agent a payoff of zero. To see this, it is sufficient to establish that no contract implementing $[s, f]$ following disclosure does as well for the principal. To implement $[s, f]$ following disclosure of $[p'(\varepsilon), 0]$, a contract $\bar{w}$ must satisfy (i) $0 \leq \bar{w}(y) \leq y$, and (ii) $E_s[\bar{w}(y)] - f \geq E_{p'(\varepsilon)}[\bar{w}(y)]$. From (ii), substituting for $p'(\varepsilon)$ yields $E_s[\bar{w}(y)] \geq f + (1+\varepsilon)(1-\beta)E_s[\bar{w}(y)]$, or equivalently $E_s[\bar{w}(y)] \geq \frac{f}{\beta-\varepsilon(1-\beta)}$. For $\varepsilon \in (0, \frac{\beta}{1-\beta})$, this implies $E_s[\bar{w}(y)] > \frac{f}{\beta} = \beta E_s[y]$, which implies $E_s[y - \bar{w}(y)] < (1-\beta)E_s[y] \leq E_{p'(\varepsilon)}[y]$, where the left-hand term is the principal's payoff from implementing $[s, f]$ using contract $\bar{w}$ and the rightmost term is the corresponding payoff under the zero contract. Thus, for any $\varepsilon \in (0, \frac{\beta}{1-\beta})$, following disclosure of $[p'(\varepsilon), 0]$ the principal strictly prefers offering the zero contract and implementing $[p'(\varepsilon), 0]$ to any other feasible contract. Therefore any such agent does not disclose, i.e., $[p'(\varepsilon), 0] \in \mathcal{N}(\beta y)$ for all $\varepsilon \in (0, \frac{\beta}{1-\beta})$. It follows by taking the limit as $\varepsilon \to 0$, that $V_P^D(\beta y) \leq (1-\beta)E_{p'(0)}[y] = (1-\beta)(E_s[y] - \frac{f}{\beta}) = V_P(\beta y)$ and therefore $V_P^D(\beta y) = V_P(\beta y)$.

It remains to show that there is no other contract $w' \in \mathbb{W}$ that is a robustly strictly profitable deviation from βy following non-disclosure. Formally, we want to show that for each $w' \neq \beta y$, $V_P^D(w'; \mathcal{N}(\beta y)) \leq V_P^D(\beta y)$. The agent $[\delta_0, 0]$ has a payoff $\beta E_s[y] - f > 0$ following non-disclosure given contract βy. Following disclosure, since $[s, f]$ is the only positive surplus action available and $[\delta_0, 0]$ is available and $w'(y) \leq y$, $[s, f]$ is the only action the principal can implement at a positive profit. Furthermore, any optimal contract that does so will leave zero payoff for the agent (e.g., the contract αy with $\alpha = \frac{f}{E_s[y]}$ is optimal). Therefore, $[\delta_0, 0] \in \mathcal{N}(\beta y)$. For any w' such that $[s, f] \notin \mathcal{A}^*(w' \mid [s, f])$, $V_P(w' \mid \mathcal{A}_0 \cup \{[\delta_0, 0]\}) \leq 0$ and therefore, $V_P^D(w'; \mathcal{N}(\beta y)) \leq V_P(w' \mid \mathcal{A}_0 \cup \{[\delta_0, 0]\}) \leq 0 < V_P^D(\beta y)$.

The only contracts w' it remains to consider are those with $[s, f] \in \mathcal{A}^*(w' \mid [s, f])$ and with $V_P(w' \mid \mathcal{A}_0) > V_P^D(\beta y)$. For each such w', let $z(w') = 1 - \frac{V_P^D(\beta y)}{V_P(w' \mid \mathcal{A}_0)}$, and consider the agents $[p(w', \varepsilon), 0]$ for $p(w', \varepsilon) = (1+\varepsilon)(1-z(w'))s + (z(w') - \varepsilon(1-z(w')))\delta_0$, which is a probability distribution for $\varepsilon \in [0, \frac{z(w')}{1-z(w')}]$. Following disclosure of $[p(w', \varepsilon), 0]$, if the principal offered the best contract implementing $[s, f]$, which we denote $\bar{w}$, then $E_s[\bar{w}(y)] - f = (1+\varepsilon)(1-z(w'))E_s[\bar{w}(y)]$ would imply profit $E_s[y] - \frac{f}{z(w')-\varepsilon(1-z(w'))}$ for the principal and payoff $\frac{f}{z(w')-\varepsilon(1-z(w'))} - f$ to the agent, and if the principal offered the zero contract and implemented $[p(w', \varepsilon), 0]$ then the profit would be $(1+\varepsilon)(1-z(w'))E_s[y]$ and the agent's payoff would be zero. When $z(w') - \varepsilon(1-z(w')) > \beta$, the principal will optimally offer $\bar{w}$ and therefore the agent is willing to not disclose because their anticipated non-disclosure payoff under βy is at least $\min\{\beta(1+\varepsilon)(1-z(w'))E_s[y], \beta E_s[y] - f\}$, which, given the inequality at the start of this sentence, is at least as large as the disclosure payoff $\frac{f}{z(w')-\varepsilon(1-z(w'))} -$

f. When $z(w') - \varepsilon(1 - z(w')) \leq \beta$, the principal will optimally offer the zero contract following disclosure and therefore the agent strictly prefers not to disclose because their anticipated non-disclosure payoff under βy is positive, i.e., $[p(w', \varepsilon), 0] \in \mathcal{N}(\beta y)$. Following non-disclosure, for $\varepsilon \in (0, \frac{z(w')}{1-z(w')})$, each agent $[p(w', \varepsilon), 0]$ takes their additional action under w'. To see this, observe that $E_{p(w',\varepsilon)}[w'(y)] = (1+\varepsilon)(1-z(w'))E_s[w'(y)] > \frac{V_P^D(\beta y)}{E_s[y-w'(y)]}E_s[w'(y)]$. Furthermore, since $E_s[w'(y)] \in [f, E_s[y]]$ and $\max_{x \in [f, E_s[y]]} \frac{(E_s[y]-x)(x-f)}{x} = (\sqrt{E_s[y]} - \sqrt{f})^2 = V_P^D(\beta y)$, $V_P^D(\beta y) \geq \frac{E_s[y-w'(y)](E_s[w'(y)]-f)}{E_s[w'(y)]}$. Therefore $E_{p(w',\varepsilon)}[w'(y)] > \frac{V_P^D(\beta y)}{E_s[y-w'(y)]}E_s[w'(y)] \geq \frac{E_s[y-w'(y)](E_s[w'(y)]-f)}{E_s[w'(y)]E_s[y-w'(y)]}E_s[w'(y)] = E_s[w'(y)] - f$. Putting everything together, and taking the limit as $\varepsilon \to 0$, $V_P^D(w'; \mathcal{N}(\beta y)) \leq E_{p(w',0)}[y - w'(y)] = \frac{V_P^D(\beta y)}{V_P(w'|\mathcal{A}_0)}E_s[y - w'(y)] = V_P^D(\beta y)$ since w' implements $[s, f]$ from $\mathcal{A}_0$. □

3.4 Preservation of robustly optimal randomized contracts under voluntary disclosure

Until now, like most of the literature, we have considered only deterministic contracts. As Kambhampati (2023) and Kambhampati et al. (2025) point out, choosing to randomize over contracts can be advantageous for the principal in the robust contracting problem. A natural question becomes whether the addition of voluntary disclosure affects the form of robustly optimal randomized contracts. We show that, in contrast to what we saw can happen without randomization, any robustly optimal randomized contract remains part of an equilibrium under voluntary disclosure in which that contract is offered following non-disclosure. Thus, robust optimality of these contracts is maintained under voluntary disclosure.

Consider the same model earlier with the following modification: The principal's set of feasible actions is expanded from $\mathbb{W}$ to $\Delta(\mathbb{W})$. Correspondingly, the principal's robust objective functions $V_P(w)$ and $V_P^D(w; \mathcal{N})$ are extended to $\Delta(\mathbb{W})$ in the natural way: choose $\omega \in \Delta(\mathbb{W})$ to maximize $V_P(\omega) = \inf_{[p,c]} E_\omega V_P(w \mid [p, c])$ and $V_P^D(\omega; \mathcal{N}) = \inf_{[p,c] \in \mathcal{N}} E_\omega V_P(w \mid [p, c])$, respectively.

Theorem 3.5. *Consider any equilibrium robust randomized contract when no disclosure is allowed. There exists an equilibrium when voluntary disclosure is allowed in which this randomized contract is offered following non-disclosure in some equilibrium.*

Proof. Let ω^* be the randomized contract offered by the principal in an equilibrium when no disclosure is allowed. Let I be the set of Borel-measurable, bounded real-valued functions on $\Delta(\mathcal{Y}) \times \mathbb{R}_+$. Observe that both the upper contour set, $D_1 \equiv \{a \in I \mid \inf_{[p,c]} a > V_P(\omega^*)\}$, and the feasible set $D_2 \equiv \{a \in I \mid a = E_\omega V_P(w \mid [p, c]), \omega \in \Delta(\mathbb{W})\}$ are convex, non-empty and disjoint. By a separating hyperplane theorem (e.g., Aliprantis and Border 1999, Thm.

5.50, p. 190), there exists a hyperplane separating them. This implies the existence of a finitely-additive measure, r, over actions such that $E_r[a] \geq E_r E_{\omega^*} V_P(w \mid [p,c]) \geq E_r[b]$ for all $a \in D_1, b \in D_2$. By standard arguments it follows that $E_\mu E_{\omega^*} V_P(w \mid [p,c]) \geq E_r E_{\omega^*} V_P(w \mid [p,c]) \geq E_r E_\omega V_P(w \mid [p,c])$ for all finitely-additive measures μ over actions and all $\omega \in \Delta(\mathbb{W})$. Therefore $E_{\omega^*} V_P(w \mid [p,c]) \geq E_{\omega^*} V_P(w \mid [p^*,c^*]) \geq E_\omega V_P(w \mid [p^*,c^*])$ for all $[p^*,c^*]$ in the support of r and all actions $[p,c]$ and all $\omega \in \Delta(\mathbb{W})$. Observe that the latter inequality says that ω^* is a best response by the principal following disclosure of $[p^*,c^*]$. When voluntary disclosure is allowed, consider a strategy profile satisfying the following: the principal offers ω^* following non-disclosure and following disclosure of $[p^*,c^*]$, and offers some randomized contract maximizing $E_\omega V_P(w \mid [p,c])$ following disclosure of any other $[p,c]$, the agent plays the same best response to any contract in the support of ω^* as in the equilibrium when no disclosure is allowed, and any best response, with ties broken in favor of the principal, when facing any other contract, and $[p^*,c^*] \in \mathcal{N}$ and the disclosure decision of all other actions are best responses given the principal's strategy and the agent's reactions to contracts. To verify that any such strategy profile is an equilibrium, first observe that the agent in the last stage indeed best responds to any contract offer faced, so that equilibrium condition (i) is satisfied. Next, $[p^*,c^*] \in \mathcal{N}$ satisfies condition (ii) since such an agent expects the principal to offer according to ω^* following both non-disclosure and disclosure. All other agents' disclosure decisions satisfy condition (ii) by construction. Finally, following non-disclosure, since $[p^*,c^*] \in \mathcal{N}$ and $E_{\omega^*} V_P(w \mid [p,c]) \geq E_{\omega^*} V_P(w \mid [p^*,c^*]) \geq E_\omega V_P(w \mid [p^*,c^*])$, $V_P(\omega^*) = E_{\omega^*} V_P(w \mid [p^*,c^*]) = V_P^D(\omega^*;\mathcal{N}) \geq V_P^D(\omega;\mathcal{N})$ for all $\omega \in \Delta(\mathbb{W})$, thus equilibrium condition (iii) is satisfied following non-disclosure, and also following any disclosure by construction and the earlier observation that ω^* is a best response by the principal following disclosure of $[p^*,c^*]$. $\square$

References

1. Acharya, V. V., DeMarzo, P., & Kremer, I. (2011). Endogenous information flows and the clustering of announcements. *American Economic Review*, 101(7), 2955-79.

2. Aliprantis, C. D., & Border K. C. (1999), Infinite Dimensional Analysis: A Hitchhiker's Guide, second edition. Springer, Berlin.

3. Antić, N. (2021). Contracting with unknown technologies. mimeo., Northwestern University.

4. Antić, N. & Georgiadis, G. (2025). Robust Contracts: A Revealed Preference Approach. *Review of Economics and Statistics*, forthcoming.

5. Ben-Porath, E., & Lipman, B. L. (2012). Implementation with partial provability. *Journal of Economic Theory*, 147(5), 1689-1724.

6. Ben-Porath, E., Dekel, E., & Lipman, B. L. (2019). Mechanisms with evidence: Commitment and robustness. *Econometrica*, 87(2), 529-566.

7. Burkett, J. & Rosenthal, M. (2024). Data-driven contract design. *Journal of Economic Theory*, 221, 105900.

8. Carroll, G. (2015). Robustness and linear contracts. *American Economic Review*, 105(2), 536-63.

9. Carroll, G. (2019). Robustness in mechanism design and contracting. *Annual Review of Economics*, 11, 139-166.

10. DeMarzo, P. M., Kremer, I., & Skrzypacz, A. (2019). Test design and minimum standards. *American Economic Review*, 109(6), 2173-2207.

11. Diamond, P. (1998). Managerial incentives: On the near linearity of optimal compensation. *Journal of Political Economy*, 106(5), 931-957.

12. Dye, R. A. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research*, 23(1), 123-145.

13. Eső, P., & Wallace, C. (2014). Information and evidence in bargaining. *Economic Theory Bulletin*, 2(1), 23-32.

14. Eső, P., & Wallace, C. (2019). Bargaining with hard evidence. *The Economic Journal*, 129(621), 2039-2063.

15. Eső, E., & Wallace, C. (2022). Evidence disclosure in competitive markets. mimeo., University of Oxford.

16. Gajdos, T., Hayashi, T., Tallon, J.-M., & Vergnaud, J.-C. (2008). Attitude toward imprecise information. *Journal of Economic Theory*, 140, 27-65.

17. Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2), 141-153.

18. Glode, V., Opp, C. C., & Zhang, X. (2018). Voluntary disclosure in bilateral transactions. *Journal of Economic Theory*, 175, 652-688.

19. Gode, D. D. K., & Singh, R. (2006). The impact of verifiability on contracts. *Journal of Accounting, Auditing & Finance*, 21(2), 149-168.

20. Hanany, E., Klibanoff, P., & Mukerji, S. (2020). Incomplete information games with ambiguity averse players. *American Economic Journal: Microeconomics*, 12(2), 135-87.

21. Innes, R. D. (1990). Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory*, 52(1), 45-67.

22. Kambhampati, A. (2023). Randomization is optimal in the robust principal-agent problem. *Journal of Economic Theory*, 207, 105585.

23. Kambhampati, A., Peng, B., Tang, Z. G., Toikka, J., Vohra, R. (2025). Randomization and the Robustness of Linear Contracts. mimeo., US Naval Academy.

24. Olszewski, W. (2025), On evaluating contracts by their performance in the worst cases. mimeo., Northwestern University.

25. Shavell, S. (1994). Acquisition and disclosure of information prior to sale. *The RAND Journal of Economics*, 25(1), 20-36.

26. Shin, H. S. (1994a). News management and the value of firms. *The RAND Journal of Economics*, 25(1), 58-71.

27. Shin, H. S. (1994b). The burden of proof in a game of persuasion. *Journal of Economic Theory*, 64(1), 253-264.

28. Shin, H. S. (2003). Disclosures and asset returns. *Econometrica*, 71(1), 105-133.

29. Walton and Carroll (2022). A general framework for robust contracting models. Econometrica 90(5), 2129–2159.