Contractual Traps

February 21, 2009

Abstract

In numerous economic scenarios, contracting parties may not have a clear picture of all the relevant aspects. While confronted with these unawareness issues, the strategic decisions of the contracting parties critically depend on their sophistication. A contracting party may be unaware of what she is entitled to determine. Therefore, she can only infer some missing pieces via the contract offered by other parties and determine whether to accept the contract based on her own evaluation of how reasonable the contract is. Further, a contracting party may actively gather information and collect evidence about all possible contingencies to avoid to be trapped into the contractual agreement. In this paper, we propose a general framework to investigate these strategic interactions with unawareness, reasoning, and cognition. We build our conceptual framework upon the classical principal-agent relationship and compare the equilibrium behaviors under various degrees of the unaware agent's sophistication. Several implications regarding optimal contract design, possible exploitation, and cognitive thinking are also presented.

Keywords: Unawareness, cognition, incomplete contracts, principal-agent relationship
JEL Classification: D86, D82, D83

"*Nature never deceives us; it is we who deceive ourselves.*"

——Jean Jacques Rousseau

1 Introduction

In numerous economic scenarios, contracting parties may not have a clear picture of all the relevant aspects. A contracting party may be *unaware* of what she herself and other contracting parties are entitled to determine. For example, an employee may be unaware of the possibility of obtaining some training to improve her productivity, and may not know ex ante that the employer could provide a poor retirement plan. A car buyer may be unaware that the dealer may secretly modify the specs of the car (e.g., whether the deal includes the air conditioning, built-in GPS, extended warranty, and rear seat entertainment system) that are not explicitly written in the contract. An insuree may be unaware that an insurer may delay or withhold the repayments of her life insurance. This unawareness issue also arises when consumers are surprised by add-on costs of cartridge after buying a printer, or the costs of using the telephone, watching in-room movies in a hotel, and so on.

While confronted with these unawareness issues and the potential exploitation by others, the strategic decisions of the contracting parties critically depend on their *sophistication*. A naive contracting party may take the contract offer as given and passively updates her view of the world through the unexpected terms specified in the contract. A more sophisticated contracting party may attempt to put herself on the others' feet to evaluate whether a proposed contract is a honest mutually beneficial deal, a sloppy mistaken contract offered by a careless partner, or a trap intentionally set up to take advantage of her. Further, a contracting party may actively gather information and collect evidence about all possible contingencies in order to compensate/ overcome the asymmetric awareness. These counteractions are all natural defensive responses that a rational contracting party can take in order to protect herself from being cheated by others, or, in our terminology, being *trapped* into a contractual relationship.

When a contractual relationship involves such unawareness, reasoning, and cognitive thinking, the optimal contract design (from the contract proposer's perspective) becomes subtle. On one hand, since the contract follower (hereafter the *agent*) is not fully aware of all the aspects relevant to the contractual relationship, the contract proposer (hereafter the *principal*) may strategically

disclose only a subset of relevant aspects in the contract at his own benefit. On the other hand, the intentionally concealed information may make a sophisticated agent suspect something may go wrong and take some defensive counteraction such as refusing the contract or actively gathering information. These inherent economic trade-offs give rise to a number of interesting issues. Given a contract offer, how does an unaware agent update her information? How does an agent rationalize the principal's contract offer? If a contract offer is unintended, how does the agent perceive and respond? How should the principal design the optimal contracts based on the agent's sophistication?

To address these issues, we construct a stylized model in which a principal intends to hire an agent to implement a project. The project requires multiple inputs of both the principal and the agent. As is standard in the principal-agent literature, we assume that all actions of the agent are not observable whereas all actions of the principal are verifiable. However, the agent may be *unaware* of all the relevant aspects she or the principal is entitled to choose. On the contrary, the principal is fully aware of the entire strategy sets of both the principal and the agent and knows the agent's awareness. Since the agent may be unaware, the principal can determine whether to inform the agent via the contract offers. This contract offer may serve as an *eye-opener* that broadens the agent's vision and allows the agent to get a better understanding of the entire picture. Moreover, the contract is not necessarily complete if it does not specify all the utility-relevant actions/obligations.

Since the contract is allowed to be incomplete, the agent can determine the actions specified in the contract accordingly and she must "choose" unconsciously the default actions that are out of her mind. The default action is chosen unconsciously based on the *rule-guided behavior* rather than her rational calculation (as in Hayek [1967] and Vanberg [2002]). Likewise, in the aspects that the agent is unaware of, she unconsciously assumes that the principal will choose the default action and attaches the hypothetical utilities to herself and the principal, respectively. The agent's conjecture of the principal's choice in the aspect she is unaware of is not based on rational expectation, but rather on her *rule-guided perception.* The detailed discussions and formal definitions of the contract, the rule-guided behavior, and the utilities are deferred to Section 2.

Based on the above framework, we propose a number of *solution concepts* that account for various degrees of the unaware agent's sophistication. As a direct extension of the classical subgame perfect Nash equilibrium to incorporate the agent's unawareness, we first introduce the *rational equilibrium* in which the agent updates her unawareness based on the principal's contract offer. The novel feature that arises from the agent's unawareness is that there is room for the principal to determine what to announce/include in the contract and which actions to implement in the

aspects not specified in the contract. Since *the principal and the agent perceive different games*, the principal's contract offer may not be optimal from the agent's viewpoint. This is in strict contrast with the standard game theory that assumes the common knowledge on the game. This discrepancy creates room for various choices of alternative solution concepts, as we elaborate below.

The above rational equilibrium implicitly assumes that the agent takes the contract offered by the principal without thinking about whether the contract is indeed optimal for the principal. Nevertheless, as demonstrated in Filiz-Ozbay [2008], Ozbay [2008], and more fundamentally Heifetz et al. [2009], an unaware agent may still be able to evaluate whether the principal's contract offer is "reasonable." Therefore, she may be reluctant to accept a contract if she believes that this contract is not the best contract (from the principal's viewpoint) among all the feasible contracts. This gives rise to the next solution concept, namely the *justifiable equilibrium*. If based on the agent's investigation, the principal should have offered an alternative contract, the agent suspects that something has gone wrong and therefore may reject the contract to avoid the potential exploitation. The idea of justifiable equilibrium is similar to that of *forward induction* in game theory, as the subsequent player also reasons the former player's motivation upon observing the former player's actions (Fudenberg and Tirole [1994]). The agent's reasoning upon receiving a contract also constitutes a refinement of what the principal is able to offer, thereby giving rise to an additional "justifiability" constraint on the principal's side. Notably, since the agent only possesses limited awareness, her own calculation regarding the principal's utility is based on her limited awareness and thus may be wrong from the principal's viewpoint.

So far we have introduced two different solution concepts. In a rational equilibrium, the agent takes the contract as given and updates her awareness passively. On the contrary, in a justifiable equilibrium, as long as the agent finds that the contract is not justifiable, she believes that the principal is setting up a trap to take advantage of her. These two solution concepts represent the two extreme reactions from the agent's side in reasoning the principal's incentive. A natural question is whether there exist other solution concepts that lie in between the two extremes. Conceivably, when confronted with an unintended (non-justifiable) contract, the agent may believe that this unintended contract simply results from *the principal's mistake* occasionally.[1]

To incorporate this type of bounded rationality into the unawareness framework, we assume

[1]Researchers have documented experimental evidence that human beings inevitably make mistakes while choosing among multiple options even if they are fully aware that some options are better than the others; see, e.g., Anderson et al. [1998], Lim and Ho [2007], McKelvey and Palfrey [1995], and Su [2008].

that when the agent faces a contract that is not justifiable, she believes that with probability $1-\rho$ it results from the principal's mistake, and with probability ρ this unintended contract is a trap set up by the principal. With these probabilities, the agent then decides whether to accept the contract based on her expected utility, which leads to a *trap-filtered equilibrium*. Note that when $\rho = 0$, the agent is extremely confident that any unintended contract should be attributed to the principal's mistake, and the trap-filtered equilibrium degenerates to a rational equilibrium. On the other hand, if $\rho = 1$, whenever she sees an unintended contract, she perceives it as a trap and the trap-filtered equilibrium coincides with the justifiable equilibrium. Thus, the trap-filtered equilibrium can be regarded as a broader family of the solution concepts and it nicely unifies all possible scenarios regarding how the agent perceives the principal's contract offer.

Finally, we investigate the scenario in which the agent is able to "think" upon receiving a non-justifiable contract. This *cognitive thinking* allows the agent to pull back from being trapped into an intentional non-justifiable contract with the principal a contract. As in Tirole [2008], such cognitive thinking is definitely helpful for the agent, but it comes at a cost. The higher cognitive effort the agent spends ex ante, the more likely she is able to identify a contractual trap. Thus, the principal must take into account the agent's cognitive thinking and the possible consequences upon designing the contract. It is worth mentioning that based on our definition of the *trap-filtered equilibrium with cognition*, the agent does not exert cognitive effort only if she sees a justifiable contract. In contrast, in Tirole [2008], the agent will not exert cognitive effort only if the principal opens the agent's eyes.[2]

Having introduced all the relevant solutions concepts, we use a stylized car-buying example to demonstrate their similarities and differences. Through this example, we observe that the principal is able to exploit the agent by offering a non-justifiable contract when the agent passively updates her unawareness, but such an exploitation becomes impossible when the agent is able to reason how the principal fares upon offering such a "too-good-to-be-true" contract. Further, if the agent may interpret the non-justifiable contract as the principal's mistake, this exploitation is more likely to occur when the contractual traps are less common. The ability of cognitive thinking allows the agent to escape from a potential contractual trap, and the agent exerts more cognitive effort when the trap is more likely to happen. Naturally, these implications should also hold in a number of economic contexts in which the contracting party suffers from the unawareness and the degree of

[2]Note also that in his framework, there is common knowledge of the game and rationality. This implies that the equilibrium contract is always justifiable. Nevertheless, cognitive thinking still occurs even though justifiability is guaranteed. Please see Section 3 for details.

sophistication is crucial.

Since we incorporate unawareness to the principal-agent relationship, our paper is related to vast literature on the unawareness. Modica and Rustichini [1994] first discuss the unawareness issue formally in economic theory, and Dekel et al. [1998] later show that it is impossible to model the non-trivial unawareness by using the standard state space. Nevertheless, Galanis [2007], Heifetz et al. [2006], and Li [2008] circumvent this negative result. The shared feature of these papers is that what is missing in the agent's mind is not arbitrary points in the state space but rather a *whole dimension* of it. We apply this idea to our contracting problems, as in Eliaz and Spiegler [2006], Filiz-Ozbay [2008], Gabaix and Laibson [2006], and von Thadden and Zhao [2007]. Our principal-agent framework extends the standard moral hazard model, see, e.g., Grossman and Hart [1983], Holmstrom [1979], Holmstrom and Milgrom [1991], and Mirrlees [1999]. Unlike the aforementioned work, we incorporate the agent's unawareness, which gives rise to the novel issue of whether the principal should propose an incomplete contract.

Our paper is also related to the literature on incomplete contracts. This literature proposes several rationales for contractual incompleteness: verifiability (Grossman and Hart [1986] and Hart and Moore [1990]), signaling (Aghion and Bolton [1987], Chung and Fortnow [2007], and Spier [1992]), explicit writing costs (Anderlini and Felli [1999], Battigalli and Maggi [2002], and Dye [1985]), and inadequate cognition (Bolton and Faure-Grimaud [2009] and Tirole [2008]). In contrast with the above papers, we interpret the contract incompleteness as a result of the principal's incentive to optimally determine the degree of the agent's unawareness. It is also worth mentioning that Tirole [2008] introduces the contract incompleteness from a very different angle. Namely, in Tirole [2008], a more complete contract implies more cognitive efforts of the agent before contracting. In contrast, in our paper, a contract is incomplete if it does not specify all the utility-relevant actions.

The remainder of this paper is organized as follows. In Section 2, we introduce the principal-agent framework, and in Section 3 we propose a number of solution concepts and discuss the equilibrium behaviors under those solution concepts. In Section 4, we demonstrate the implications of these solution concepts in an example. Section 5 concludes, discusses some variants of our model characteristics, and evaluates the robustness and the limitations of our results.

2 The Model

We consider a stylized model in which a principal (P) intends to hire an agent (A) to implement a project. The project requires the inputs of both the principal and the agent. Specifically, let S_P and S_A denote the sets of strategies of the principal and the agent, respectively. To incorporate the possibility that each party may determine multiple decisions, we allow S_P and S_A to include multiple components (dimensions): $S_P \equiv A_P^1 \times \ldots \times A_P^M$ and $S_A \equiv A_A^1 \times \ldots \times A_A^N$ with M, $N < \infty$. In the canonical employee compensation example, the employer (the principal) may determine the compensation scheme that comprises the fixed payment and the commission rate for the employee (the agent). The employer may further determine other actions such as the employee's retirement benefit. These decisions affect directly the utilities of the employer and the agent and are included in S_P. On the employee's side, she may have the discretion of determining how much effort to exert in completing the project or whether to receive some external training that improves her productivity.

We use $s_P \equiv (a_P^1, \ldots, a_P^M)$ and $s_A \equiv (a_A^1, \ldots, a_A^N)$ to denote the elements in the strategy sets of the principal and the agent, respectively. Further, let $S \equiv S_P \times S_A$ with $s \in S$. To avoid the technical difficulties, we assume that the strategy (action) space, S, is finite.[3] Given the strategy profiles s_P and s_A, the principal and the agent obtain utilities u_P and u_A, respectively, where $u_i : S \mapsto \mathbb{R}$, $i \in \{P, A\}$, is a mapping from the entire strategy profiles (of both the principal and the agent) to a real-valued utility. If eventually the project is not implemented, they receive the reservation utilities $\overline{u}_P$ and $\overline{u}_A$ that correspond to the utilities they obtain from their outside options.

In contrast with the standard principal-agent models, we assume that the agent may be unaware of all the relevant aspects she or the principal is entitled to choose. Specifically, let $D_i \equiv \{A_i^1, A_i^2, \ldots\}$ denote the collection of all action sets of party i, and $D \equiv D_P \cup D_A$ denotes the collection of all action sets of both the principal and the agent. Let W_i ($W_i \subseteq D_i$) denote the set of action sets of i of which the agent is aware *before* contracting, where $i \in \{P, A\}$. Thus, $W \equiv W_P \cup W_A$ represents the collection of action sets that the agent is aware of. On the contrary, we assume that the principal is fully aware of both the entire strategy set S and the agent's

[3] In this way, the games (to be defined later) will be finite as well. The existence of equilibrium (under the appropriate solution concepts) can be easily established following the classical game theory literature (see, e.g., Fudenberg and Tirole [1994]).

awareness (i.e., W).[4] In this sense, the principal is omniscient: he knows the entire picture of the economic context, and he knows precisely what is endowed in the agent's mind.[5]

Since the agent may be unaware, the principal can determine whether to inform the agent via the contract offers. This contract offer may serve as an *eye-opener* that broadens the agent's vision and allows the agent to get a better understanding of the entire picture. Obviously, the principal must indicate in the contract the corresponding compensations based on all the actions that the agent is aware of (i.e., W); additionally, the principal might announce actions that are out of the agent's mind. Specifically, let $V = V_P \cup V_A$ represent the collection of action sets that are specified in the contract but are out of the agent's mind. We can interpret V as the principal's strategic announcement to alleviate the agent's unawareness.

Contract. We can now formally define a contract offered by the principal. In the following, we use $\times X$ to denote the Cartesian product of all action sets in $X \subseteq D$, i.e., $\times X \equiv \Pi_{Y \in X} Y$.

Definition 1. *A **contract** is a vector $\psi(V) \in \times(W \cup V)$ where $V \subseteq D \setminus W$.*[6]

Note that ψ specifies the compensation the agent collects given the agent's action in $W \cup V$. Let $\psi(V) \equiv (\psi_P(V), \psi_A(V))$ where $\psi_i(V)$ is composed only of party i's actions. Following the literature that incorporates the unawareness into the contracting framework, we assume that whenever the principal announces some actions that are out of the agent's mind, the agent is able to understand the contract immediately and adjust her awareness to account for the additional aspects specified in the contract; see, e.g., Filiz-Ozbay [2008] and Ozbay [2008].

We can now define the contract completeness based on the above notion:

Definition 2. *A contract $\psi(V)$ is **incomplete** in party i's strategy if $W_i \cup V_i \neq D_i$, where $i \in \{P, A\}$.*

By definition, a contract is incomplete in party i's strategy if it does not specify the complete

[4]The situation where the principal is uncertain about the agent's awareness and therefore screens the agent's awareness is studied by Eliaz and Spiegler [2006] and von Thadden and Zhao [2007].

[5]It is possible to extend our analysis to the case in which the principal is only partially aware following the approach developed in Zhao [2008]. Since our focus is on the impact of the agent's sophistication on the optimal contract design, we exclude the possibility of the principal's awareness.

[6]The order of the elements is based on the following rule: The action sets of the principal precede the action sets of the agent and $\forall i$, A_i^k precedes A_i^l if and only if $k < l$. For example, if $W \cup V = \{A_A^2, A_P^3, A_P^1\}$, then $\times W \equiv A_P^1 \times A_P^3 \times A_A^2$.

utility-relevant actions that party i can select.[7] We say a contract ψ is *incomplete* if ψ is incomplete in either the principal's or the agent's strategy. Given a contract $\psi(V)$, the agent's effective strategy, denoted by $s_A(V)$, is confined within $\times(W \cup V)$; likewise, $s_P(V) \in \times(W_P \cup V_P)$ corresponds to a feasible strategy profile for the principal *from the agent's perspective*. In general, $s(V) \equiv (s_P(V), s_A(V))$ is an incomplete strategy profile, since it is composed of the actions only in the agent's mind. The larger the set V is, the more dimensions the vector $s(V)$ has.

Although an incomplete contract does not specify the complete utility-relevant actions/obligations, it provides clear instructions of actions in some dimensions ($W_i \cup V_i$ for party i). If the actions are observable and are written in the contract, they are perfectly enforceable. Moreover, only these actions are enforced. In the legal language, this corresponds to the extreme legal environment in which there is no mandatory and default rules on each dimension of parties' actions. The role of the court is passive in that it treats a written contract as complete and thus forbids all extrinsic evidence to clarify the ambiguity in the contract on the unspecified dimensions of actions.[8]

Rule-guided behavior. Since the contract is allowed to be incomplete, if $\psi(V)$ is incomplete in the agent's strategy, the agent can determine the actions specified in the contract accordingly and she must "choose" *unconsciously* the actions that are out of her mind. In this paper, we assume that if the agent is unaware of some aspect $A_A^k \notin W_A \cup V_A$ after observing the contract, she unconsciously choose her *default action* $\bar{a}_A^k$ in this aspect. This default action is assumed to be unique for simplicity. Likewise, for $A_P^k \notin W_P \cup V_P$, the agent unconsciously assumes that the principal will choose the default action $\bar{a}_P^k$ (which is assumed to be unique as well). Extensions to the scenarios where there are multiple default actions are straightforward.

Let us elaborate more on the interpretation of the default actions. As the agent is unaware of A_A^k, the default action $\bar{a}_A^k$ is chosen unconsciously based on her *rule-guided behavior* rather than her rational calculation. As in Hayek [1967] and Vanberg [2002], the rule-guided behavior is orthogonal to the conscious process; the rule simply decodes the contractual situation facing the agent and gives an instruction $\bar{a}_A^k$ to the agent. Since this rule is completely out of the agent's mind, the agent simply follows the rule without even noticing it. As an example, in the employee compensation problem, if an employee is unaware of the possibility of obtaining some training to improve her

[7] It is worth mentioning that Tirole [2008] interprets the contract completeness from a very different angle. Namely, he argues that the contract is more complete if the agent exerts more cognitive efforts before contracting. In contrast, in our paper, a contract is incomplete if it does not specify all the complete payoff-relevant actions.

[8] This coincides with the Willistonian or "textualist" approach, which argues that the contract is the only document that the court can use to determine the plain meaning of the contracting parties, see Schwartz and Scott [2003].

productivity, she may simply ignore the training without any contemplation. In such a scenario, receiving no training is her default action in this aspect.

The agent's unawareness is also reflected in how she perceives what the principal would do and how her own utility is affected. If the agent is unaware of A_P^k (i.e., $A_P^k \notin W_P \cup V_P$), the agent unconsciously takes for granted that the principal should choose $\bar{a}_P^k$ and, unconsciously, takes this default action $\bar{a}_P^k$ into her own utility function. In this sense, the agent's conjecture of the principal's choice in the aspect she is unaware of is not based on rational expectation, but rather on her *rule-guided perception.* This rule-guided perception can be regarded as a hypothesis in the agent's mind. The agent is unaware of this hypothesis in her mind even after the contracting stage; moreover, this hypothesis could be wrong. In the example of the employee compensation problem, if the employee is unaware that her employer could provide a poor retirement plan, then the employee may contemplate whether to accept the contract as if the retirement plan would be not that bad if she believes so. The employee's decision is based on this hypothesis, which may be wrong if ex post the employer indeed provides a poor retirement plan.

In general, let us denote $s^C(V) \equiv (s_P^C(V), s_A^C(V)) \in \times(D \setminus (W \cup V))$ as the set of actions that the agent is unaware of. The complete (objective) strategy profile $s = (s(V), s^C(V))$ is composed of both the strategy profiles in and out of the agent's mind. If the principal indeed chooses the default action in the aspects that the agent is unaware of, the strategy profile then satisfies that $s_P^C(V)$ consists of only default actions $\bar{a}_P^k$. Define $\overline{s}(V) \equiv (\overline{s}_P(V), \overline{s}_A(V))$ as this special case. Note that the principal has the discretion to choose any feasible action in the aspect out of the agent's mind. Thus, the principal's effective strategy space expands to the entire S_P. For example, if the obligation of an employer in the contract is only to fulfill the compensation level, then nothing prevents the employer from offering a low retirement benefit, or postponing the salary payment.

Subjective utilities. Given the aforementioned rule-guided behavior and the agent's unawareness, we can then articulate how the agent evaluates a contract $\psi(V)$. Let $u_i^V : \times(W \cup V) \mapsto \mathbb{R}$, $i \in \{P, A\}$, denote the subjective utility function of party i *from the agent's viewpoint.*[9] From the representation, the function u_i^V clearly depends on the strategy space V specified in the contract

[9]The term "subjective" refers to the subjectivity of belief but not the subjectivity of preference. Notably, belief-subjective utility could be wrong, since the individual "believes" in a specific utility form that is merely a hypothesis in her mind. On the other hand, since preference-subjective utility corresponds to the individual's true "feeling" (that represents his personal value judgment), this utility is always correct.

(and the corresponding actions $s(V)$). In the presence of the agent's unawareness, we assume that

$$u_i^V(\cdot) = u_i(\cdot, \bar{s}(V)), i \in \{P, A\},$$

for consistency, where $u_i : S \mapsto \mathbb{R}$ is the *objective* utility function of party i if every aspect is known. This reflects that the subjective utility functions $\{u_i^V(\cdot)\}$'s are consistent with the objective utility functions $u_i(\cdot)$ where the missing variables are completed by the default strategy profile $\bar{s}(V)$. Thus, the agent simply believes that the default actions will be taken in the aspects she is unaware of, and derives the corresponding (subjective) utilities for herself and the principal. Notably, we can conveniently incorporate some uncertain components into the utility functions.[10]

As is standard in the principal-agent literature, we assume that all actions of the agent are not observable whereas all actions of the principal are verifiable.[11] Furthermore, we assume that the principal always intends to have the project implemented as opposed to opting for his outside option. A sufficient but crude condition is that $\inf_{s \in S} u_P(s) \geq \overline{u}_P$, where $\overline{u}_P$ corresponds to the principal's reservation utility as aforementioned. Nevertheless, the agent may be better off to turn down the contract offer. Specifically, if we define $\inf_{s \in S} u_A(s)$ as the agent's worst-case utility level if she accepts the contract, this implies that $\inf_{s \in S} u_A(s) < \overline{u}_A$. This assumption is adopted in the remainder of this paper. As we demonstrate later, this assumption simply rules out the trivial case in which the agent always accepts the contract even if the principal may deceive her (because the agent's outside option is extremely unfavorable).

Having described the contract, the rule-guided behavior, and the utilities, we in the next section investigate how the principal and the agent should make their decisions regarding the contract offer and the actions.

[10] In such a scenario, we can regard the utility u_i as the expected utility over all possible contingencies. Specifically, if $\tilde{u}_i(\cdot, \epsilon)$ denotes the realized utility of player i, where ϵ captures the residual uncertainty, then $u_i(\cdot) \equiv E_\epsilon \tilde{u}_i(\cdot, \epsilon)$ is the effective utility that depends only on the selection of actions. Since we focus on the unawareness of the actions rather than the unawareness of the states, the probability distribution of the residual uncertainty ϵ should be common knowledge.

[11] This may not be appropriate in certain scenarios, but modifications are straightforward. On the agent's side, contractible actions can be directly written into the contract and thus unawareness does not matter. On the principal's side, if the agent is unaware of a specific action of the principal, it makes no difference whether this action is observable or not. If the agent is aware of a principal's action but this is non-contractible, the contract should also provide an incentive scheme for the principal to induce the appropriate action choice as in the double moral hazard problems.

3 Equilibrium Analysis

In this section, we provide predictions of the "equilibrium" behaviors of the principal and the agent. To this end, it is essential to define what decision rules should the principal and the agent follow in equilibrium. In the terminology of game theory, these rules are described by the "*solution concepts*" (see Fudenberg and Tirole [1994]). In the standard moral hazard model in which every aspect is known to both parties, we can conveniently adopt subgame perfect Nash equilibrium as the solution concept. Since the game involves the agent's unawareness, subgame perfect Nash equilibrium is no longer appropriate. In the following, we first provide some preliminary discussions of the essential components of the equilibrium behavior, and then introduce a number of solution concepts that are suitable for the economic environments that involve unawareness.

3.1 Preliminaries

Before we introduce the solution concepts, it is helpful to specify the timing in this contractual relationship. The sequence of events is as follows. 1) The principal proposes the contract $\psi(V)$; upon observing the contract, the agent updates her awareness. 2) The agent decides whether to accept the contract. If not, the game is over and both parties receive their reservation utilities from the outside options. 3) If the contract is accepted, the agent chooses s_A and unconsciously implements $\overline{s}_A(V)$ in $s_A^C(V)$; the principal chooses s_P afterwards.

We now introduce some definitions regarding a contract offer. Since there might be discrepancy between the principal's claimed actions and the realized actions, we define $(\psi(V), s)$ as an *assessment*. Given the contract and the agent's updated unawareness, we can describe how the agent makes the contract choice and whether to accept the contract or not. As in the standard principal-agent problems, the equilibrium action (contract) choice of the agent must be *incentive compatible* (IC):

$$\psi_A(V) \in \arg\max_{\widetilde{\psi}_A(V)} u_A^V(\psi_P(V), \widetilde{\psi}_A(V)), \tag{IC}$$

where $u_A^V(\psi_P(V), \widetilde{\psi}_A(V))$ in the right-hand side is the agent's subjective utility under a specific strategy profile $\widetilde{\psi}_A(V)$ and the belief that the principal chooses $\psi_P(V)$. The incentive compatibility constraint guarantees that the equilibrium action profile maximizes the agent's (subjective) utility.

Furthermore, in order to induce the agent to accept the contract, the following *individual*

rationality (IR) constraint should hold:

$$u_A^V(\psi(V)) \geq \overline{u}_A. \tag{IR}$$

We can now define the set of feasible contracts.

Definition 3. *A contract $\psi(V)$ is **feasible** if it satisfies (IC) and (IR).*

The above discussions concern whether the agent is willing to follow the equilibrium behavior. On the other hand, in order to sustain an equilibrium, the principal should not be better off via any possible deviation, which gives rise to the following definition.

Definition 4. *An assessment $(\psi(V), s)$ is **consistent** if $\psi(V) = s(V)$ and $s_A^C(V) = \overline{s}_A(V)$.*

The consistency of an assessment ensures that the principal's observed actions are the same as his claimed actions in the contract (and the agent chooses the default actions in the dimensions he is not aware of). Feasibility and consistency are maintained throughout this paper in every solution concept, as we describe next.

3.2 Rational Equilibrium

The first solution concept is the rational equilibrium, which essentially follows from the solution concept in the standard principal-agent problem.

Definition 5. *An assessment $(\psi^*(V^*), s^*)$ is a **rational equilibrium** if*

- $(V^*, \psi^*(V^*), s^*) \in \arg\max\, u_P$,
- $\psi^*(V^*)$ *is feasible,*
- $(\psi^*(V^*), s^*)$ *is consistent.*

To interpret the rational equilibrium, it is helpful to first review the procedure to obtain the solution to a standard principal-agent problem. Without the issue of unawareness, this is done in two stages. In the first stage, the contract must satisfy the incentive compatibility and individual rationality constraints (or collectively the feasibility). In the second stage, among the set of feasible contracts, the principal must select the one that maximizes his (expected) utility. When the agent is unaware of some aspects, there is room for the principal to determine what to announce/include

in the contract and which actions to implement in the aspects not specified in the contract. Thus, to deter any deviation from the principal, in addition to the feasibility constraint, the principal faces two more incentive constraints: 1) given the contract, the principal must implement the actions as expected (the consistency issue); and 2) the information conveyed in the contract must be optimal from the principal's perspective. Note that since the principal strictly prefers to have the agent participate, no trade is never an equilibrium. Further, as A_i^k is finite for all i and k, the rational equilibrium exists because the game is finite as well (see, e.g., Fudenberg and Tirole [1994]).

The rational equilibrium can be regarded as a direct extension of the classical subgame perfect Nash equilibrium to incorporate the agent's unawareness. Recall that in a subgame perfect Nash equilibrium, at each node of the game, the active player should update her belief given the past history, and based on the updated belief, her equilibrium strategy must maximize her utility from then on (Fudenberg and Tirole [1994]). Due to this subgame perfect feature, the game is solved by *backward induction* starting from each end node of the game. As we apply the subgame perfect Nash equilibrium to our context, we shall first focus on the agent's problem. Here, the novel feature is the agent's unawareness. Thus, similar to the belief updating in the subgame perfect Nash equilibrium, the agent in our model must update her unawareness based on the principal's contract offer. The principal perfectly foresees the agent's response and then optimally determines the contract offer (and which set of actions to include in the contract).

However, because the principal and the agent perceive different games (due to the agent's unawareness), the principal's contract offer may not be optimal from the agent's viewpoint. This is in strict contrast with the standard game theory that assumes the common knowledge on the game. This discrepancy creates room for various choices of alternative solution concepts, as we elaborate in the subsequent sections.

3.3 Justifiable Equilibrium

In the rational equilibrium, a critical assumption is that the agent takes the contract offered by the principal without thinking about whether the contract is indeed optimal for the principal. This does not cause any problem if the agent were fully aware of all the aspects. Nevertheless, as demonstrated in Filiz-Ozbay [2008] and Ozbay [2008], an unaware agent may be reluctant to accept a contract if she believes that this contract is not the best contract (from the principal's viewpoint) among all the feasible contracts. This gives rise to the next solution concept, namely

the justifiable equilibrium.

Before introducing the solution concept, let us first define a justifiable contract.

Definition 6. *A contract $\psi(V)$ is **justifiable** if*

- *it is feasible;*
- $\forall \widetilde{V} \subseteq V$, $\forall \psi(\widetilde{V}) \in \times(W \cup \widetilde{V})$, $\widetilde{s}(V) \in \times(W \cup V)$ *such that $\psi(\widetilde{V})$ is feasible and $(\psi(\widetilde{V}), \widetilde{s})$ is consistent, we have $u_P^V(\psi(V)) \geq u_P^V(\widetilde{s}(V))$.*

According to the above definition, a contract is justifiable if the agent thinks that the principal indeed proposes an optimal contract. Note that since this can only be verified after the agent considers every possible contract that the principal would propose, an implicit assumption is that the agent is aware that she may be unaware of something. This assumption is also adopted in Filiz-Ozbay [2008] and Ozbay [2008]. Moreover, from the definition of a justifiable contract, the agent takes into consideration her own best response for every given contract. Thus, she believes that the principal can perfectly predict how the agent would behave (in the sense of rational equilibrium). All the above descriptions require a significant amount of rationality and reasoning. Notably, since the agent only possesses limited awareness, her own calculation regarding the principal's utility is based on the subjective utility (u_P^V) rather than the objective utility (u_P). Thus, this may be wrong from the principal's viewpoint.

When the agent is able to think about that the principal indeed offers his optimal contract, her participation decision critically depends on whether the principal's contract offer is "reasonable." If based on the agent's investigation, the principal should have offered an alternative contract, the agent then suspects that something has gone wrong and therefore feels deceived due to her unawareness. In such a scenario, whether the agent should accept the contract or not is determined by what utility she attaches to the contract. As the principal may offer an unintended contract, an extremely "ambiguity averse" agent may assume the *worst case* scenario upon accepting the contract, which gives rise to the lowest utility $\inf_{s \in S} u_A(s)$. Since we assume that $\inf_{s \in S} u_A(s) \leq \overline{u}_A$, the agent should reject the contract. Of course, the agent might not obtain $\inf_{s \in S} u_A(s)$ when the contract is indeed a trap. However, since the agent does not know what the trap is – or at least the agent is unable to predict how the principal would behave given an unintended contract offer, it is convenient to assume that in the agent's mind, a contractual trap leads to the worst-case utility $\inf_{s \in S} u_A(s)$.

The modified sequence of events is as follows. 1) The principal proposes the contract $\psi(V)$; 2) The agent evaluates whether the contract is indeed the best interest of the principal; if not, she rejects the contract immediately; 3) After the agent's evaluation, if the contract is also optimal for the principal, the agent decides whether to accept the contract. 4) If the contract is rejected, both parties obtain their outside options; if it is accepted, the agent chooses s_A and unconsciously implements $\overline{s}_A(V)$ in $s_A^C(V)$; the principal then chooses s_P.

We next define the justifiable equilibrium.

Definition 7. *An assessment* $(\psi^*(V^*), s^*)$ *is a* ***justifiable equilibrium*** *if*

- $(V^*, \psi^*(V^*), s^*) \in \arg\max\, u_P$;
- $\psi^*(V^*)$ *is justifiable;*
- $(\psi^*(V^*), s^*)$ *is consistent.*

In a justifiable equilibrium, we impose, on top of the standard incentive compatibility constraints, the justifiability constraint on the principal's side. As the key difference between the rational and justifiable equilibria, this justifiability ensures that the principal offers the contract that is optimal for him based on the agent's calculation, and it significantly restricts the principal's choice of contract in order to induce the agent's participation.[12] The existence of a justifiable equilibrium can be easily established.[13]

The idea of justifiable equilibrium is similar to that of *forward induction* in game theory, as the subsequent player also reasons the former player's motivation upon observing the former player's actions. Recall that forward induction requires each player to rationalize other players' behaviors and actively interpret the rationale for an unintended action (Fudenberg and Tirole [1994]). In

[12] A subtle point in the justifiability is that in the agent's mind, the principal believes that the agent simply gives a best response to the contract (within the agent's awareness). In other words, the agent believes that the principal is unaware that the agent can evaluate the justifiability of the contract. This occurs in every subgame in which the principal offers a non-justifiable contract. However, on the equilibrium path, the principal takes it into consideration (and offers a justifiable contract); therefore, the agent has a wrong belief regarding the principal's sophistication; see also Ozbay [2008] for the discussions on justifiability in a different context.

[13] The idea goes as follows. The principal can simply choose the best contract (among those feasible contracts) that is justifiable for him. Given this contract, the agent accepts the contract and then determines her optimal actions accordingly. Note also that there exists at least one justifiable contract: the one that makes the agent fully aware, albeit it may be suboptimal.

our context, since the principal is omniscient but the agent is not fully aware, the idea of forward induction applies naturally to the agent rather than the principal. The agent's reasoning upon receiving a contract also constitutes a refinement of what the principal is able to offer. Moreover, this refinement is extremely restrictive in that any contract is rejected by the agent as long as it does not qualify to be justifiable. The agent's unwillingness to accept an unintended contract follows from our assumption that $\inf_{s \in S} u_A(s) < \overline{u}_A$.[14]

So far we have introduced two different solution concepts. In a rational equilibrium, the agent takes the contract as given and updates her awareness passively. On the contrary, in a justifiable equilibrium, the agent rejects the contract whenever she thinks the principal does not offer the contract that is in the principal's best interest. These two solution concepts represent the two extreme reactions from the agent's side in reasoning the principal's incentive. A natural question is whether there exist other solution concepts that lie in between the two extremes. This motivates us to propose the next solution concept.

3.4 Trap-filtered Equilibrium

In the above discussions, we assume that as long as the agent finds that the contract is not justifiable, she believes that the principal is setting up a trap to take advantage of her, thereby rejecting the contract immediately. In this sense, from the agent's perspective, the principal is fully unreliable; on the other hand, the agent completely trusts the principal's rationality. This may appear to be a strong assumption in some scenarios. For example, it is possible that the agent believes that this unintended contract simply results from *the principal's mistake*. Researchers have documented experimental evidence that human beings inevitably make mistakes while choosing among multiple options even if they are fully aware that some options are better than others. A growing stream of literature relates this to the "future uncertainty" and uses this to explain the "trembling-hand" behavior widely observed in the experiments; see the *quantal response equilibrium* literature such as Anderson et al. [1998], Lim and Ho [2007], McKelvey and Palfrey [1995], and Su [2008].

Our goal in this section is to incorporate this type of bounded rationality into the unawareness framework. To focus on our primary interest – the unawareness issue, we abstract away from the detailed probability calculations proposed by the literature on quantal response equilibrium. Instead, we simply assume that when the agent faces a contract $\psi(V)$ that is not justifiable,

[14] If this assumption is violated, i.e., $\inf_{s \in S} u_A(s) > \overline{u}_A$, the rational equilibrium suffices to be the appropriate solution concept even if the agent is more sophisticated. In this sense, the forward induction step becomes unnecessary.

she simply believes that with probability $1-\rho$ it results from the principal's mistake, and with probability ρ this contract is a trap set up by the principal.

One way to interpret this is that from the agent's viewpoint, there are two types of principals: a rational one and an irrational one. The irrational principal always makes a mistake (i.e., offering a non-justifiable contract). However, a rational principal may intentionally set up a trap for the agent. There are also two types of games: the one where the agent knows the game and the one the agent is unaware of something (and thus does not know the actual game). The agent is uncertain of the principal's rationality and her knowledge of the game, and the values of these two dimensions are independent. The probabilistic system is lexicographic in the sense of Blume et al. [1991]: the event that the principal is rational (irrational) is first-order (second-order), and the event that the agent knows (is unaware of) the game is first-order (second-order). Thus, if the agent faces a justifiable contract, the agent believes that with probability one she knows the game and the principal is rational. However, conditional on a non-justifiable contract, the agent believes that there is a trap with probability ρ.

With these probabilities, we can then express the agent's expected utility upon observing the contract $\psi(V)$ as follows:

$$U_A^T(\rho, \psi(V)) := \rho \inf_{s \in S} u_A(s) + (1-\rho) u_A^V(\psi(V)),$$

where $\inf_{s \in S} u_A(s)$ corresponds to the agent's worst-case utility if she believes that the contract is a trap, and $u_A^V(\psi(V))$ is the agent's utility after she updates her awareness and chooses the optimal strategies accordingly. Given the agent's belief about the principal's behavior, the agent accepts the contract $\psi(V)$ if the following individual rationality constraint is satisfied:

$$U_A^T(\rho, \psi(V)) \geq \overline{u}_A. \tag{IR-T}$$

We can now define an acceptable contract when the agent believes in the possibility of the principal's mistake and the corresponding solution concept.

Definition 8. *A contract $\psi(V)$ is **trap-filtered** if 1) it is justifiable or 2) it is feasible and (IR-T) is satisfied.*

The idea behind the above definition is that the agent believes that the principal may cheat her only if the contract is not justifiable. In such a scenario, a non-justifiable contract makes the agent suspect whether it is indeed in the best interest of the principal, thereby giving rise to the

second set of condition. Note that neither condition is implied by the other: It is possible that a justifiable contract does not satisfy (IR-T), and a contract that satisfies condition (2) need not be justifiable.

The next step gives a formal definition of a trap-filtered equilibrium.

Definition 9. *An assessment* $(\psi^*(V^*), s^*)$ *is a* ***trap-filtered equilibrium*** *if*

- $(V^*, \psi^*(V^*), s^*) \in \arg\max\, u_P$;
- $\psi^*(V^*)$ *is trap-filtered;*
- $(\psi^*(V^*), s^*)$ *is consistent.*

Note that when $\rho = 0$, the agent is extremely confident that any unintended contract should be attributed to the principal's mistake; she proceeds to update her awareness according to the contract and determines her optimal strategies, and the trap-filtered equilibrium degenerates to a rational equilibrium. On the other hand, if $\rho = 1$, the agent believes that the principal never makes a mistake; thus, whenever she sees an unintended contract, she perceives it as a trap and the trap-filtered equilibrium coincides with the justifiable equilibrium. Thus, the trap-filtered equilibrium can be regarded as a broader family of the solution concepts that incorporate the ones reported in the literature. Equilibrium existence follows the similar arguments and therefore is omitted. In Section 4, we illustrate the differences and similarities of these solution concepts via an example.

3.5 Trap-filtered Equilibrium with Cognition

The trap-filtered equilibrium has nicely unified all possible scenarios regarding how the agent perceives the principal's contract offer. Nevertheless, in all the aforementioned solution concepts, the agent can only passively interpret the principal's behavior and react accordingly based on her conservativeness and confidence. While this might be satisfactory in certain scenarios, it could also be possible that the agent is able to "think" through the scenarios upon receiving a contract. Of course, if the contract offer is justifiable, such *cognitive thinking* does not benefit the agent, since there is no trap with probability one due to the lexicographic probabilistic system by Blume et al. [1991] as we discussed earlier; however, if the principal indeed offers a non-justifiable contract, thinking allows the agent to pull back from being trapped into a contract. As in Tirole [2008], such cognitive thinking is typically costly and the associated cost is implicit and frequently ignored

in the classical contract theory. Our goal, in this section, is to incorporate the cognition into our contractual framework with unawareness.

To formalize our ideas, we assume that the agent can spend some cost in evaluating whether a non-justifiable contract is due to the principal's mistake or the agent's unawareness. This cognition stage arises after the principal has offered the contract but before the agent decides whether to accept the contract. The higher cost the agent spends ex ante, the more likely she is able to identify a contractual trap given that there is indeed a trap. Specifically, let $c \in [0, 1]$ denote the probability that the agent finds out that the contract is a trap (conditional on the event that it is indeed a trap). The associated cost of cognitive thinking is denoted by an increasing function $T(c)$. Note that even though the agent actively thinks through the scenarios, it is still possible that the principal may trap the agent via a non-justifiable contract (but less likely due to the agent's cognitive effort).

With the addition of the cognition stage, the modified sequence of events is as follows. 1) The principal proposes the contract $\psi(V)$. 2) Upon receiving the contract, the agent (costlessly) evaluates whether the contract is justifiable. 3) If the contract is justifiable, the agent spends no cognitive cost and determines directly whether to accept the contract; if the contract is non-justifiable, the agent makes the cognitive thinking and evaluates whether the contract is a trap or simply a principal's mistake. 4) After the cognition stage, if the agent figures out that a non-justifiable contract is a trap, she refuses to sign a contract and the game ends immediately; if based on her cognitive thinking, the agent thinks it is more likely to be the principal's mistake, she then determines whether to accept the contract.[15] 5) Finally, if the contract is accepted, the principal and the agent make their decisions and obtain their utilities.

Let us articulate how the agent decides whether to accept the contract. Suppose that ex ante the agent decides to spend the cognitive thinking cost $T(c)$. Upon observing a non-justifiable contract, with probability $1-\rho$, the agent believes that this comes entirely from the principal's mistake and therefore proceeds to update her unawareness. In this case, she obtains utility $u_A^V(\psi(V))$ upon accepting the contract. With probability ρc, the agent figures out that the contract is an intentional trap. To be consistent with the scenarios discussed earlier, we assume that the agent rejects the contract if she thinks it is a trap and obtains her reservation utility $\overline{u}_A$. Finally, with probability

[15] This is different from Tirole [2008] in that cognitive thinking occurs when the contract is not justifiable, whereas in Tirole [2008], the agent exerts cognitive thinking only when the agent's eyes are not opened. In essence, a non-justifiable contract and non eye-opening information play the same role in the situations where something may go wrong for the agent. However, our formulation allows the possibility of seeing a "too-good-to-be-true" contract that would never occur in Tirole [2008].

$\rho(1-c)$, the agent cannot figure out the trap and attaches ex ante the utility $\inf_{s\in S} u_A(s)$ to such an event. Collectively, the agent's ex ante expected utility is

$$U_A^C(c,\rho,\psi(V)) \equiv \rho c\overline{u}_A + \rho(1-c)\inf_{s\in S} u_A(s) + (1-\rho)u_A^V(\psi(V)) - T(c).$$

This determines the optimal cognitive cost spending as follows:[16]

$$c^*(\rho,\psi(V)) \in \arg\max_{c\in[0,1]} U_A^C(c,\rho,\psi(V)),$$

and the corresponding optimal expected utility is

$$U_A^C(\rho,\psi(V)) \equiv U_A^C(c^*(\rho,\psi(V)),\rho,\psi(V)).$$

The agent will accept a non-justifiable contract $\psi(V)$ if and only if the following ex ante individual rationality constraint holds:

$$U_A^C(\rho,\psi(V)) \geq \overline{u}_A. \tag{IR-C}$$

Note also that cognitive thinking allows the agent to determine whether to accept the contract *after* the cognition stage.[17]

We can now introduce the solution concept with cognition.

Definition 10. *A contract $\psi(V)$ is **trap-filtered with cognition** if it is justifiable, or it is feasible and (IR-C) holds.*

Definition 11. *An assessment $(\psi^*(V^*), s^*, c^*)$ is a **trap-filtered equilibrium with cognition** if*

- $(V^*, \psi^*(V^*), s^*) \in \arg\max \{c^*\overline{u}_p + (1-c^*)u_P(s)\}$;
- *$\psi^*(V^*)$ is trap-filtered with cognition;*

[16]Note that in this formulation, the cognitive effort can take value from a continuous support $[0,1]$. This implies that the game is no longer finite, and the argument for establishing the equilibrium existence does not apply. However, a finite game is sufficient for existence but not necessary. As we demonstrate in Section 4, a trap-filtered equilibrium with cognition may still exist even if the strategy space is not finite.

[17]We can also write down the ex post individual rationality constraint (after the cognition stage). It requires that

$$\frac{\rho(1-c)}{\rho(1-c)+1-\rho}\inf_{s\in S} u_A(s) + \frac{1-\rho}{\rho(1-c)+1-\rho}u_A^V(\psi(V)) \geq \overline{u}_A, \tag{IR-C2}$$

where $\rho(1-c)+1-\rho$ is the probability that the agent does not find any evidence of the contractual trap, and $\frac{\rho(1-c)}{\rho(1-c)+1-\rho}$ and $\frac{1-\rho}{\rho(1-c)+1-\rho}$ are the conditional probabilities that a non-justifiable contract is indeed a trap or a result of the principal's mistake, respectively. In fact, (IR-C) implies (IR-C2) because (IR-C) implicitly assumes that the agent will accept the contract ex post.

- $(\psi^*(V^*), s^*)$ *is consistent;*
- $c^* = 0$ *if* $\psi^*(V^*)$ *is justifiable and* $c^* \in \arg\max_{c\in[0,1]} U_A^C(c, \rho, \psi(V))$ *otherwise.*

Recall that there are two types of principals: a rational one and an irrational one, and a rational principal may intentionally set up a trap for the agent. Thus, in the definition of the trap-filtered equilibrium with cognition, the rational principal intends to choose $(V^*, \psi^*(V^*), s^*)$ to maximize

$$c^*\overline{u}_p + (1 - c^*)u_P(s),$$

where the term $c^*\overline{u}_p$ corresponds to the case in which the cognitive thinking is effective (which occurs with probability c^*), and the second term corresponds to the case in which the agent accepts the contract and makes the optimal actions accordingly. In response to the potential contractual trap from the rational principal, the agent exerts the optimal cognitive effort to figure out whether there is a contractual trap. Upon receiving a non-justifiable contract, in the agent's mind, there is distinction between two cases: 1) The principal makes a mistake (which occurs with probability $1 - \rho$); and 2) The principal indeed sets up a trap but the agent fails to catch it (with probability $\rho(1 - c)$).

We have introduced a sequence of solution concepts that assume different degrees of rationality and cognitive ability on the agent. In the next section, we provide various examples to demonstrate the similarities and differences of these solution concepts.

4 A Numerical Example

In this section, we demonstrate the differences among these solution concepts via a numerical example. In this example, both parties have two dimensions of strategies. The first dimension action set A_i^1 is a singleton $\left\{\overline{a}_i^1\right\}$ which consists of a usual action of the party i. The second dimension of actions is however out of the agent's mind. For simplicity, let us assume that $A_A^2 = \{0, 1\}$ and $A_P^2 = \{0, 2\}$, and the default actions are $\overline{a}_P^2 = \overline{a}_A^2 = 0$. The alternative actions $a_P^2 = 2$ and $a_A^2 = 1$ are the unforeseen actions for the agent. In our notation, $W = \left\{A_P^1, A_A^1\right\}$ since the agent is only aware of the usual actions of both parties in the first dimension.

To visualize this example, suppose that a principal intends to sell a car to an agent.[18] We

[18]Here, rather than giving a typical employment example in the standard principal-agent relationship, we choose to focus on a buyer-seller relationship to demonstrate the flexibility of our model.

can interpret the first dimension as the typical reception from the principal as the agent enters the store. In the second dimension, the agent's choice (if she is aware) is between a status quo car and a novel car, and the principal's corresponding action is whether to provide the air conditioning in the car. The agent's default action ($\overline{a}_A^2 = 0$) is to choose a status quo car and the principal's default action ($\overline{a}_P^2 = 0$) is to provide the air conditioning. Further, assume that the principal must provide the air conditioning in the status quo car, but he is able to remove it from the novel car.[19] The alternative action $a_A^2 = 1$ corresponds to the case in which the agent chooses a novel car, and $a_P^2 = 2$ corresponds to the principal's decision to remove the air conditioning from the novel car. This saves the principal's cost but reduces the agent's utility upon purchasing.

Given the two dimensions of actions, the objective utilities of the principal and the agent are respectively $u_P = a_P^1 a_A^1 - a_A^2 + a_A^2 a_P^2$ and $u_A = a_P^1 a_A^1 + a_A^2 - a_A^2 a_P^2$. Since A_i^1 is a singleton, we can conveniently assume that the default (regular) actions are both 1 (i.e., $\overline{a}_P^1 = \overline{a}_A^1 = 1$). After these substitutions, we obtain that $u_P = 1 - a_A^2 + a_A^2 a_P^2$ and $u_A = 1 + a_A^2 - a_A^2 a_P^2$. Let the reservation utilities of them are $\overline{u}_P = \delta$ and $\overline{u}_A = 1$ (which correspond to the situation in which no trade occurs). Note that in order to guarantee that the principal always intends to induce the agent's participation, we require that $\delta < 0$.

Let us first consider the scenario in which the principal does not announce any new actions (the option of buying a novel car) to the agent, i.e., $V = \emptyset$. In such a scenario, the agent can only decide between purchasing the status quo car ($a_A^2 = 0$) and simply walking away. Given this, since the principal cannot remove the air conditioning, the principal 's action affects neither the agent nor the principal himself. Therefore, choosing $a_P^2 = 0$ is the principal's best response, and as a result both the principal and the agent obtain utility 1.

If, on the contrary, the principal informs the agent of the possibility of choosing the novel car (i.e, $V = \left\{A_A^2\right\}$), the agent is then aware of this new dimension and therefore makes the decision optimally based on her subjective utility. In this case, since the principal does not disclose his own action set A_P^2 (that he may remove the air conditioning), under the solution concept of rational equilibrium, the agent continues to (unconsciously) believe that the principal will provide the air conditioning ($a_P^2 = 0$). Thus, from the agent's perspective, her subjective utility is $u_A^{A_A^2} = 1 + a_A^2$. The corresponding best response is to choose $a_A^2 = 1$ and in the agent's mind she should obtain a

[19] This assumption ensures that if the principal intends to set up a trap, he can only do so upon introducing the novel car. If the principal is also allowed to remove the air conditioning secretly from a status quo car, the trap could appear in all scenarios.

subjective utility 2.

We now turn to the principal's problem. By backward induction, the principal perfectly foresees the agent's action $a_A^2 = 1$. Consequently, his (objective) utility becomes $u_P = a_P^2$ and thus his optimal strategy is to choose $a_P^2 = 2$. From the above discussion,

$$(\psi(V), s) = ((\overline{a}_P^1, \overline{a}_A^1, 1), (\overline{a}_P^1, 2, \overline{a}_A^1, 1))$$

is the unique rational equilibrium. The principal proposes the novel car for the agent, but does not mention the possibility of removing the air conditioning. Notably, this equilibrium gives rise to a utility 2 for the principal but an actual utility $1 + a_A^2 - a_A^2 a_P^2 = 0$ for the agent, whereas in the agent's mind the supposed utility is 2 rather than 0. In this sense, the contract $\psi(V)$ with $V = \{A_A^2\}$ is a trap for the agent. The agent takes the lure of the novel car and thus is willing to choose $a_A^2 = 1$. The principal then takes advantage from the agent by removing the air conditioning ($a_P^2 = 2$).

The above discussions demonstrate how a contractual trap can be implemented even if the agent is fully rational (but is subject to her unawareness). We next apply the idea of justifiability to this example. When the agent is sophisticated, she may feel that the novel car is "*too good to be true*." This is because in the agent's mind, if A_A^2 were not specified in the contract, the principal would receive utility $u_P^{A_A^2} = 1$. However, the contract with $V = \{A_A^2\}$ offers the agent an opportunity to choose an action a_A^2 which benefits the agent herself but might hurt the principal as the principal receives utility $u_P^{A_A^2} = 0$. Thus the contract in the rational equilibrium is *not justifiable*. Note that from the agent's perspective, the principal's utility crucially depends on the offered contract due to the agent's updated awareness. When $V = \emptyset$, the agent believes that $u_P = 1$; when $V = \{A_A^2\}$, it becomes $u_P = 1 - a_A^2$; when $V = \{A_P^2,\ A_A^2\}$, the subjective utility becomes $u_P = 1 - a_A^2 + a_A^2 a_P^2$.

Next, we assume that the agent believes that the non-justifiable contract (regarding the novel car) may result from the principal's mistake (with probability $1-\rho$). It follows from straightforward algebra that the contract with $V = \{A_A^2\}$ is a trap-filtered equilibrium when $\rho \leq \frac{1}{2}$. When the probability of the principal's mistake is high, upon receiving a non-justifiable contract, the agent is more inclined to interpret it as a mistake and consequently accepts the contract with $V = \{A_A^2\}$ although it is too good to be true. In other words, the principal sets a trap only when the agent believes that the non-justifiability of the contract is more likely due to the principal's mistake rather than a trap. This coincides with our intuition: In a society where contractual traps are not

common, the agent is more inclined to accept non-justifiable contracts. On the principal's side, the rational principal is (weakly) better off for a higher ρ as the trap is easier to implement, because the principal may receive utility $u_P = 2$ rather than $u_P = 1$ for a smaller ρ.

Finally, we introduce the cognitive thinking. For simplicity let $T(c) = \frac{1}{2}c^2$. Following from the definition of the trap-filtered equilibrium with cognition, the agent accepts the contract only if

$$\max_{c \geq 0} \left\{ \rho c + 2(1-\rho) - \frac{1}{2}c^2 \right\} \geq 1,$$

that is, $\rho \leq 2 - \sqrt{2} \approx 0.58579$. Since the agent's ability to conduct cognitive thinking allows her to reject a contract after the cognition stage, it is conceivable that she can afford to accept the contract more likely (i.e., with a higher ρ compared to the case without the cognitive thinking) and the agent should obtain a higher expected utility. We further find that in equilibrium $c = \rho$, i.e., the more likely there is a trap, the more effort the agent exerts in the cognitive thinking.

In the presence of cognition stage, the principal sets up a trap only if $2(1-c) + c\delta \geq 1$, that is, $\rho(2-\delta) \leq 1$. This also has an intuitive interpretation. When the principal's outside utility (δ) is low, he is severely punished by the agent's non-participation once the contractual trap is caught. Thus, the principal's incentive to set a contractual trap declines as δ becomes lower. As in the case without cognition stage, we also observe that the possibility of setting a contractual trap is higher when the agent is more convinced that this results from the principal's mistake (ρ is low). Notably, the rational principal may be weakly better off when the agent is endowed with the ability to conduct cognitive thinking because the condition for the agent to accept the contract is weaker (as $0.58579 > \frac{1}{2}$).

To summarize, if the agent is endowed with the ability of cognitive thinking, $(\psi(V), s) = ((\overline{a}_P^1, \overline{a}_A^1, 1), (\overline{a}_P^1, 2, \overline{a}_A^1, 1))$ is a trap-filtered equilibrium with cognition if $\rho(2-\delta) \leq 1$ and $\rho \leq 0.58579$. Note that in this example, the support of the cognitive effort is continuous rather than finite. However, a trap-filtered equilibrium with cognition still exists.

In this example, we observe that the principal is able to exploit the agent by offering a non-justifiable contract when the agent passively updates her unawareness, but such an exploitation becomes impossible when the agent is able to reason how the principal fares upon offering such a "too-good-to-be-true" contract. Further, if the agent may interpret the non-justifiable contract as a principal's mistake, this exploitation is more likely to occur when the contractual traps are less common. The ability of cognitive thinking allows the agent to escape from a potential contractual trap, and the agent exerts more cognitive effort when the trap is more likely to happen.

5 Concluding Remarks

In this paper, we propose a general framework to investigate these strategic interactions with the aforementioned unawareness, reasoning, and cognition. We build our conceptual framework upon the classical principal-agent relationship and compare the equilibrium behaviors under various degrees of the unaware agent's sophistication. Specifically, the rational equilibrium can be regarded as a direct extension of the classical subgame perfect Nash equilibrium. The justifiable equilibrium constitutes a refinement of what the principal is able to offer and is in line with the forward induction in game theory. The trap-filtered equilibrium incorporates the possibility that an unintended contract may simply result from the principal's mistake, and the trap-filtered equilibrium with cognition allows the agent to conduct cognitive thinking and pull back from being trapped into an intentional non-justifiable contract with the principal a contract. These solution concepts are well suited in various economic contexts that involve the contracting parties' unawareness, bounded rationality, psychological effect, and cognition.

The primary message we intend to convey in this paper is to demonstrate the possibility of incorporating unawareness, reasoning, and cognition in a unified framework. This general framework certainly has its own limitations; however, due to its simplicity, we open up a number of possible extensions for other economic contexts of interest. For example, we abstract away from the renegotiation problem in the post-contracting stage. Nevertheless, when the agent figures out that the principal's contract is non-justifiable, it is conceivable that the two contracting parties may attempt to renegotiate. The principal may intend to offer an alternative contract that takes into account the agent's updated unawareness; furthermore, the agent may also make a counter-offer to the principal. Detailed procedure of the renegotiation stage may vary depending on the relative bargaining power and the institutional convention. In such a scenario, alternative solution concepts may be proposed following the approach in Tirole [2008], and it would be intriguing to see whether this renegotiation stage influences the agent's response to the contract offer and how the principal designs the optimal contract.

Our focus on the monopolistic principal's optimal contract design problem may be a bit excessive. In certain situations, it is possible that multiple principals, either homogeneous or heterogeneous in terms of their awareness and preferences, may compete in hiring the agent that is exposed to the unawareness issue. Thus, the agent's awareness in the post-contractual stage is jointly determined by the contracts offered by these principals with conflicting interests. Another

possible extension is to introduce multiple agents with heterogeneous degrees/dimensions of unawareness. The interesting question in this alternative setting is whether the principal intends to offer secret/private contracts to these agents, and if so, whether the agents have an incentive to communicate with each other after receiving the principal's offers.

In this paper, we focus on the one-shot transaction between the principal and the agent. However, in many practical situations, these contracting parties may interact in multiple rounds. While extended to the multiple-round (repeated) setting, the optimal contract design in this principal-agent relationship becomes more sophisticated. It has been well-documented that in a dynamic contracting environment, the ratchet effect and the commitment problem significantly complicate the optimal contract design. In our framework, we impose, on top of those difficulties, the additional strategic concerns of how much information to disclose through the contract offers over time, and how much information the agent is able to infer/reason/think about given the sequence of proposed contracts. Finally, given the principal's incentive to offer the contractual trap and the agent's (wasteful) effort on cognitive thinking, it might be welfare improving if a benevolent third party is introduced to control the information flow. While our results certainly provide some preliminary policy implications for the public announcements, thorough studies on the social welfare, efficiency, and fairness are needed in order to provide a general picture, and are left for future research.

References

P. Aghion and P. Bolton. Contracts as a barrier to entry. *American Economic Review*, 77(3): 388–401, June 1987.

L. Anderlini and L. Felli. Incomplete contracts and complexity costs. *Theory and Decision*, 46: 23–50, 1999.

S. Anderson, J. Goeree, and C. Holt. Rent seeking with bounded rationality: An analysis of the all-pay auction. *Journal of Political Economy*, 106(4):828–853, 1998.

P. Battigalli and G. Maggi. Rigidity, discretion, and the costs of writing contracts. *American Economic Review*, 92(4):798–817, September 2002.

L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and equilibrium refinements. *Econometrica*, 59(1):81–98, January 1991.

P. Bolton and A. Faure-Grimaud. Satisficing contracts. NBER Working Papers 14654, National Bureau of Economic Research, Inc, January 2009.

K. Chung and L. Fortnow. Loopholes. Working paper, University of Minnesota and Northwestern University, 2007.

E. Dekel, B. Lipman, and A. Rustichini. Standard state-space models preclude unawareness. *Econometrica*, 66(1):159–174, 1998.

R. Dye. Costly contract contingencies. *International Economic Review*, 26(1):233–250, 1985.

K. Eliaz and R. Spiegler. Contracting with diversely naive agents. *Review of Economic Studies*, 73 (3):689–714, 07 2006.

E. Filiz-Ozbay. Incorporating unawareness into contract theory. Working paper, University of Maryland, 2008.

D. Fudenberg and J. Tirole. *Game Theory*. The MIT Press, Cambridge, Massachussetts, 1994.

X. Gabaix and D. Laibson. Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics*, 121(2):505–540, May 2006.

S. Galanis. Unawareness of theorems. Discussion Paper Series In Economics And Econometrics, Economics Division, School of Social Sciences, University of Southampton, 2007.

S. Grossman and O. Hart. An analysis of the principal-agent problem. *Econometrica*, 51(1):7–45, 1983.

S. Grossman and O. Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719, August 1986.

O. Hart and J. Moore. Property rights and the nature of the firm. *Journal of Political Economy*, 98(6):1119–58, December 1990.

F. Hayek. *Rules, Perception and Intelligibility*, pages 43–65. Studies in Philosophy, Politics and Economics. The University of Chicago Press, 1967.

A. Heifetz, M. Meier, and B. Schipper. Interactive unawareness. *Journal of Economic Theory*, 130 (1):78–94, September 2006.

A. Heifetz, M. Meier, and B. Schipper. Dynamic unawareness and rationalizable behavior. Working paper, The Open University of Israel, 2009.

B. Holmstrom. Moral hazard and observability. *Bell Journal of Economics*, 10(1):74–91, 1979.

B. Holmstrom and P. Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization*, 7(0):24–52, Special I 1991.

J. Li. Information structures with unawareness. Forthcoming in Journal of Economic Theory, 2008.

N. Lim and T. Ho. Designing price contracts for boundedly rational customers: Does the number of blocks matter? *Marketing Science*, 26(3):312–326, 2007.

R. McKelvey and T. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995.

J. Mirrlees. The theory of moral hazard and unobservable behaviour: Part I. *Review of Economic Studies*, 66(1):3–21, 1999.

S. Modica and A. Rustichini. Awareness and partitional information structures. *Theory and Decision*, 37(1):107–124, 1994.

E. Ozbay. Unawareness and strategic announcements in games with uncertainty. Working paper, University of Maryland, 2008.

A. Schwartz and R. Scott. Contract theory and the limits of contract law. *Yale Law Journal*, 113: 541–619, 2003.

K. Spier. Incomplete contracts and signalling. *RAND Journal of Economics*, 23(3):432–443, Autumn 1992.

X. Su. Bounded rationality in newsvendor models. Forthcoming in Manufacturing & Service Operations Management, 2008.

J. Tirole. Cognition and incomplete contracts. Forthcoming in American Economic Review, 2008.

V. Vanberg. Rational choice vs. program-based behavior: Alternative theoretical approaches and their relevance for the study of institution. *Rationality and Society*, 14:7–53, 2002.

E. von Thadden and X. Zhao. Incentives for unaware agents. Working paper, University of Mannheim, 2007.

X. Zhao. Moral hazard with unawareness. *Rationality and Society*, 20:471–496, 2008.