Unawareness and Contracts

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Xiaojian Zhao

Mannheim 2010

Abteilungssprecher:	Prof. Tom Krebs, PhD
Referent:	Prof. Dr. Ernst-Ludwig von Thadden
Korreferent:	Prof. Dr. Martin Peitz
Tag der mündlichen Prüfung:	25nd March 2010

Acknowledgments

I am greatly indebted to my supervisor Ernst-Ludwig von Thadden for his invaluable guidance throughout my doctoral studies. I would like to thank his continuous advice on my research projects. His vision supported me in each crucial step in my academic development. Without him, I could not imagine how my life would be. I am also very grateful to Jean Tirole for his local supervision and hospitality in Toulouse school of economics. He enhanced my courage to pursue my dreams in the intellectual world, especially in social sciences.

I express my gratitude to my coauthors, Ying-Ju Chen in UC Berkeley, Roberta Dessi and Sanxi Li in Toulouse school of economics, and Martin Peitz and Ernst-Ludwig von Thadden in University of Mannheim. My research output is a result of collective efforts and would not exist without their talent and generosity.

I thank our doctoral program in economics (CDSE) and the European Network for Training in Economic Research (ENTER) for providing me a large number of opportunities of conferences, seminars, summer schools and visits. Financial support from the German Science Foundation (DFG) is gratefully acknowledged.

Furthermore, I benefited substantially from discussions with some other colleagues in University of Mannheim, University College London, Toulouse school of economics, etc. In particular, I am grateful to Syngjoo Choi, Kim-Sau Chung, Johannes Koenen, Patrick Rey, Burkhard Schipper, Heiner Schumacher, Rani Spiegler, and especially Jidong Zhou.

Lastly, my wife Qianying Chen provided invaluable support, understanding and love in my life in Europe.

Contents

Chapter 1

Introduction

Contract theory and economics of information have provided the most successful analytic framework in economic theory in the last thirty years. This paradigm helps us understand a numerous variety of social problems, say economics, finance, law, management, politics, and psychology. It raises the explanatory power of economic theory to a large extent in which classic general equilibrium theory is incompetent. Standard models in this area include screening, signaling, informed principal theory, mechanism design, cheap-talk games, information disclosure, information acquisition, information sharing, information aggregation, moral hazard, collusion, and their extensions to multi-dimensional or multi-lateral contracting environments, or dynamics. (See, e.g., Laffont and Martimort (2002) and Bolton and Dewatripont (2005).)

Nevertheless, contract theory on its own has limitations. Grossman and Hart (1986) and Hart and Moore (1990) have introduced an alternative paradigm called incomplete contracting, which provides an analytic tool to the problem of institutional design. The basic idea goes as follows. Because some contingencies (or actions) are not verifiable, contracting parties have to rely on allocation of control rights to solve the under-investment problems. However, Maskin (1999), Moore and Repullo (1988) and Maskin and Tirole (1999) use mechanism design approach to show the irrelevance of verifiability problems under some conditions. It is suggested that incomplete contracts are as a result of bounded rationality of contracting parties. However, it is easier said than done. What we need is a formal and elaborate model of bounded rationality and incomplete contracts.

All standard models in contract theory are based on standard game theory, in which most solution concepts need the agents' (high-order) knowledge of rationality and the game. In standard game theory, an agent i chooses his action s_i to maximize

$$E_{\omega\in\Omega}[u_i(\omega, s_i, s_{-i})]$$

where u_i is the agent's utility function, s_{-i} is the others' actions, and ω is some payoff-relevant contingency for the agents. As a matter of fact, there are many unknown unknowns in our everyday lives. If a person is *unaware* of something, then he does not know it, and he does not know that he does not know it, and so on *ad infinitum*.[1] Introspection tells us that unawareness is a common kind of ignorance of humans. However, it is not well-understood in the intellectual world.

In Chapter 2, we discuss the foundation of unawareness. At present, there are two parallel ways to approach unawareness: the epistemic approach and the decision theoretic approach. The epistemic approach starts from agent's belief. We construct models of knowledge in the agent's mind directly. One of the sub-approach initiated by Aumann (1976) is set-theoretic, or semantic, which is frequently used in economics. The other approach is syntactic, which is less popular in economics. In contrast, the decision theoretic approach starts from the agent's choice behaviors without referring to the agent's belief.

Notably, the thesis focuses on contracting problems with asymmetric awareness between two contracting parties. Using theoretical models, it studies the behavioral and welfare implications of unawareness in the market (and within organizations).

First, we discuss unforeseen contingencies and incomplete contracts in Chapter 3. In terms of games, the agent may not be able to perceive perfectly the entire set of contingencies ω in the future. For example, people were unaware that the earth is not flat but actually round long time ago. In the industrial revolution period, people were unaware of the future global warming. In the economic context, there are many unforeseen events made by *nature*. For example, consider that an insuree buys some home insurance. His probability judgment of the event "calamity" may be smaller than the judgment of the event "fire, explosion, earthquake, lightning, theft, storm or flood, $\cdots$", since not all these states of nature are available in the insuree's mind when only the general term "calamity" is mentioned. In this chapter, we model unawareness of contingencies and its economic implication by non-Bayesian and Bayesian approaches respectively.

Second, we examine unawareness of actions in Chapter 4. The agent is probably not able to perceive perfectly the other agent's actions (s_{-i}), or his own choice possibilities (s_i). Several years ago, the U.S. people were unaware of September 11, 2001 attacks. In the economy, many unforeseen events are directly *man-made*, that is, an economic agent's surprise is the result of the actions of the other agents. For example, 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

[1] This is the key difference between unawareness and the other kind of ignorance commonly discussed in economics, namely uncertainty. If people are uncertain of something, they know that they do not know it.

after buying a printer, or the costs of using the telephone, watching in-room movies in a hotel, and so on. Further, the agent may be unaware of his own choice possibilities. For instance, an employee may be unaware of the possibility of obtaining some training to improve his productivity, or might be unaware of some shirking behavior, such as idling about in "Second Life" in his office, that impairs his performance. In this chapter, we model contractual traps, and design the optimal incentive contracts for possibly unaware agents.

Chapter 5 concludes by discussing some relevant issues, say self-awareness and intrapersonal contracts.

In particular, the thesis covers five papers in my research portfolio. Li, Peitz and Zhao (2009) is in Section 3.1.1. Zhao (2009a) is in Section 3.2. Zhao (2009b) is in Section 3.3.2. Chen and Zhao (2009) is in Section 4.1. von Thadden and Zhao (2009) is in Section 4.3.

Chapter 2

Foundations of Unawareness

2.1 Epistemic Foundations

In this section, we use epistemic approach to model agent's belief in economics where we focus only on the semantic approach.

2.1.1 Knowledge

Before introducing unawareness, we have a short review of knowledge theory, based on which unawareness is developed. An *information structure* is a collection $\mathcal{L} \equiv (\Omega, (\mathcal{F}_i, p_i)_{i \in N})$. Ω is the set of finite states of nature. Each agent $i \in N$, where N is finite set of agents, has a *possibility correspondence* $\mathcal{F}_i : \Omega \longmapsto 2^{\Omega}$, where $\mathcal{F}_i(\omega)$ is the set of states i considers possible when ω occurs. We study two properties of $\mathcal{F}_i$.

P1 (reflexive or non deluded) $\omega \in \mathcal{F}_i(\omega)$ for all ω.

It means that i, at least, knows that the true state is possible.

P0 (consistency) For all ω, ω', we have $\omega' \in \mathcal{F}_i(\omega) \Rightarrow \mathcal{F}_i(\omega') = \mathcal{F}_i(\omega)$.

It says that i can use this consistency argument to infer about the state.

Slightly abusing notations, $\mathcal{F}_i \equiv \{F_i \subseteq \Omega : F_i = \mathcal{F}_i(\omega) \text{ for some } \omega \in \Omega\}$.

Theorem 2.1 *$\mathcal{F}_i$ is partitional if and only if $\mathcal{F}_i$ it satisfies [P1] and [P0].*

Proof 2.1 *It is clear that "only if" part directly follows. Suppose $\mathcal{F}_i$ satisfies [P1] and [P0]. If $\mathcal{F}_i(\omega')$ and $\mathcal{F}_i(\omega)$ intersect and $\omega'' \in \mathcal{F}_i(\omega') \cap \mathcal{F}_i(\omega)$, by [P0], $\mathcal{F}_i(\omega') = \mathcal{F}_i(\omega) = \mathcal{F}_i(\omega'')$. (no overlaps) By [P1], $\cup_{\omega \in \Omega} \mathcal{F}_i(\omega) = \Omega$. (no gaps) Therefore, $\mathcal{F}_i$ is partitional.*

We say i *knows* an event $A \subseteq \Omega$ at ω, if $\mathcal{F}_i(\omega) \subseteq A$. The set of states, or event, in which i knows A is $K_i(A) \equiv \{\omega \in \Omega : \mathcal{F}_i(\omega) \subseteq A\}$. $K_i(A)$ per se is an event. The knowledge operator derived from any possibility correspondence satisfies the following properties.

N: $K_i(\Omega) = \Omega$.

It says i knows something must happen. Since $\mathcal{F}_i(\omega) \subseteq \Omega$ for all ω, obviously, $K_i(\Omega) = \{\omega \in \Omega : \mathcal{F}_i(\omega) \subseteq \Omega\} = \Omega$.

MC: $K_i(A) \cap K_i(B) = K_i(A \cap B)$.

It says knowing A and knowing B is equivalent to knowing A and B. Obviously, $K_i(A) \cap K_i(B) = \{\omega : \mathcal{F}_i(\omega) \subseteq A\} \cap \{\omega : \mathcal{F}_i(\omega) \subseteq B\} = \{\omega : \mathcal{F}_i(\omega) \subseteq A \cap B\} = K_i(A \cap B)$.

Monotonicity: $A \subseteq B \Rightarrow K_i(A) \subseteq K_i(B)$.

It says if A implies B, then knowing A implies knowing B.

Theorem 2.2 (MC) *implies [Monotonicity].*

Proof 2.2 *Suppose $A \subseteq B$. $A \cap B = A$ which implies $K_i(A \cap B) = K_i(A)$. By [**MC**], $K_i(A) \cap K_i(B) = K_i(A \cap B)$. Thus $K_i(A) \cap K_i(B) = K_i(A)$. Hence $K_i(A) \subseteq K_i(B)$.*

Assume that $\mathcal{F}_i$ is partitional for all $i \in N$. We have the following properties of the knowledge operator:

T (Axiom of Knowledge): $K_i(A) \subseteq A$.

It says if i knows A, then A is true. If $\omega \in K_i(A)$, then $\mathcal{F}_i(\omega) \subseteq A$. By [P1], we have $\omega \in \mathcal{F}_i(\omega) \subseteq A$.

4 (Axiom of Transparency): $K_i(A) \subseteq K_i(K_i(A))$.

It says if i knows A, then i knows that i knows A. $K_i(A)$ is a union of elements in the partition $\mathcal{F}_i$. Since, for all F which is a union of some elements in the partition, we have $K_i(F) = F$, [**4**] follows.

5 (Axiom of Wisdom): $\neg K_i(A) \subseteq K_i(\neg K_i(A))$ where $\neg$ denotes the complement.

It says if i does not know A, then i knows that i does not know A. Since $K_i(A)$ is a union of elements in the partition $\mathcal{F}_i$, $\neg K_i(A)$ is as well, and [**5**] follows.

Note that if [**T**] is satisfied, [**4**] and [**5**] hold with equality. [1]

It is worth mentioning the following mathematical property:

Theorem 2.3 $(\Omega$, *Image*$(K_i))$ *is a topological space.*

[1]The knowledge system satisfying the above axioms is a system $S5$ in model logic.

Proof 2.3 *By [**N**], we have that Ω is in Image(K_i). Since $\mathcal{F}_i$ is a partition, by definition of K_i, we have that $\emptyset \in$Image(K_i). By [**MC**], we have that the intersection of any two elements in Image(K_i) is also in Image(K_i).*

*It is left to show that the arbitrary union $A = \cup_{\alpha \in I} A_\alpha$ is in Image(K_i) for all $A_\alpha \in$Image(K_i). To show it, let $A_\alpha = K_i(B)$ for some B. By [**4**], $K_i(A_\alpha) = K_i(K_i(B)) \supseteq K_i(B) = A_\alpha$. By [**T**], $K_i(A_\alpha) = A_\alpha$. By [**MC**], $K_i(A_\alpha) = K_i(A_\alpha \cap A) = K_i(A_\alpha) \cap K_i(A) \subseteq K_i(A)$. Thus $A_\alpha \subseteq K_i(A)$ for all A_α, which implies $A \subseteq K_i(A)$. By [**T**], $K_i(A) = A$.*

Starting from $\mathcal{F}_i$, we derive the properties of K_i. Alternatively, suppose an arbitrary set operator $K_i : 2^\Omega \longmapsto 2^\Omega$ satisfies the properties above. We define $\mathcal{F}_i : \Omega \longmapsto 2^\Omega$ by $\mathcal{F}_i(\omega) = \cap\{A \subseteq \Omega : \omega \in K_i(A)\}$. Then this $\mathcal{F}_i$ can generate the same K_i derived from $\mathcal{F}_i$. Furthermore, K_i can generate the same $\mathcal{F}_i$ derived from K_i. Note that only axioms [**MC**, **N**] are necessary and sufficient condition for the existence of a possibility correspondence. If, additionally, [**T, 4, 5**] are satisfied, the derived $\mathcal{F}_i$ is partitional.

Assume $\mathcal{F}_i$ is partitional. We say that an event A is *self-evident* to i if $\mathcal{F}_i(\omega) \subseteq A$ for all $\omega \in A$. In words, A happens if and only if i knows it happens. In terms of knowledge operator, A is self-evident to i if $K_i(A) = A$. Put it differently, A is a fixed point of the knowledge operator K_i. Finally, let R be the meet of the partitions $\mathcal{F}_i$ for all $i \in N$. That is, R is their finest common coarsening. Denote by $R(\omega)$ the element in R that contains ω. Note that $R(\omega)$ is the smallest self-evident event containing ω for all i.

We denote by $\mathcal{F}_i(A) = \cup_{\omega \in A} \mathcal{F}_i(\omega)$ the set of states which i thinks possible if some state in A occurs. Thus, at ω, i knows j knows that A occurs iff $\mathcal{F}_j(\mathcal{F}_i(\omega)) \subseteq A$, or equivalently, $\omega \in K_i(K_j(A))$.

We denote by $K_M(A)$ the event A is *mutually known* among $M \subseteq N$ where $K_M(A) \equiv \cap_{i \in M} K_i(A)$. The event A is *common knowledge* among N is $CK(A) \equiv \cap_{n=1}^{\infty} K_N^n(A)$ where K_N^n denotes n iterations of the operator K_N. Alternatively, it is equivalent to define that A is common knowledge among N at ω if $R(\omega) \subseteq A$.

2.1.2 Unawareness

Unawareness is the most commonly cited argument for dropping [**5**]. Since people often have not a complete description of the state of the world.

In this section, we drop the subscripts for all operators, since there is no interactive reasoning here.

Standard State-Space Precludes Unawareness

The argument that standard state-space precludes unawareness is based on Dekel *et. al.* (1998a). We do not define the unawareness operator from knowledge operator, but allow for an arbitrary operator satisfying certain axioms.

Plausibility: For all $A \subseteq \Omega$, $U(A) \subseteq \neg K(A) \cap \neg K(\neg K(A))$.

If an agent is unaware of something, he does not know it and he does not know that he does not know it.

KU introspection: For all $A \subseteq \Omega$, $KU(A) = \emptyset$.

It is impossible that an agent knows that he is unaware of some specific event.

AU introspection: For all $A \subseteq \Omega$, $U(A) \subseteq UU(A)$.

If an agent is unaware of some event, he is unaware that that he is unaware of this specific event.

Theorem 2.4 *Assume U satisfies [Plausibility, KU and AU introspection],*

*(i)If K satisfies [**N**], then, for all $A \subseteq \Omega$, $U(A) = \emptyset$.*

*(ii)If K satisfies [**MC**], then, for all $A, B \subseteq \Omega$, $U(A) \subseteq \neg K(B)$.*

Proof 2.4 *By [AU introspection, Plausibility], we have, for all $A \subseteq \Omega$, $U(A) \subseteq UU(A) \subseteq \neg K(\neg K(U(A)))$.*

By [KU introspection], we have, $\neg K(U(A)) = \Omega$.

Thus $U(A) \subseteq \neg K(\Omega)$.

*(i) If K satisfies [**N**], then $\neg K(\Omega) = \emptyset$. Thus $U(A) = \emptyset$.*

*(ii) If K satisfies [**MC**], which implies [Monotonicity], then, for all $B \subseteq \Omega$, $K(B) \subseteq K(\Omega)$ which implies $\neg K(\Omega) \subseteq \neg K(B)$. Thus $U(A) \subseteq \neg K(B)$.*

In words, *(i)* says the agent is aware of nothing. *(ii)* says if the agent is aware of something, he knows nothing. Since [**N**, **MC**] is the necessary condition that the knowledge operator can be derived from a possibility correspondence, *we cannot model non-trivial unawareness in the standard state-space.*

Product Models

The negative result above is solved independently by Li (2006) and Heifetz *et. al.* (2006). The shared feature of their model is that if the agent is unaware of something, what is missing in the agent's mind is not arbitrary points in the standard state space but a whole dimension of it. For example, a worker decides whether or not to sign an employment contract. There is only one aspect of the employment situation in the worker's mind: the compensation scheme. However, the quality of the working environment provided by the employer is also relevant for the worker. If the worker is unaware of the working environment when the worker signs the contract, he has to bear a bad working condition in the post-contractual stage.

Formally we regard the set of payoff relevant uncertainties is a set of questions. Let $D^* = \{D_i\}_{i \in Q}$ where D_i is the set of answers to question i and Q is the index set for all relevant questions. The full state space is $\Omega^* = \prod_{i \in Q} D_i \equiv \times D^*$. The awareness mapping is $W^* : \Omega^* \mapsto 2^{D^*}$ which assigns a subset of D^* to each full state ω^*. At full state ω^*, the subjective state space of the individual is $\Omega(\omega^*) = \times W^*(\omega^*)$. Thus the individual's subjective state space is incomplete with respect to the objective state space in the sense of missing "dimensions"(questions), but not arbitrary "points" in the objective state space.

In the employment example, for simplicity, the payoff relevant questions of the worker is $Q = \{1, 2\}$. Question 1 represents whether the salary is high, so $D_1 = \{High, Low\}$. Question 2 represents whether the working condition is good, so $D_2 = \{Good, Bad\}$. Thus the objective state space is $\Omega^* = \{(High, Good)\ , (High, Bad),\ (Low, Good),\ (Low, Bad)\}$. If the worker is provided by a good working condition, he is still unaware of it, that is, $W^*(\omega^*) = \{D_1\}$ for all $\omega^* = (\cdot, Good)$. Thus his subjective state space is $\Omega((\cdot, Good)) = \{High, Low\}$. If the worker is provided by a bad working condition, then he is suddenly aware of it, that is, $W^*(\omega^*) = D^*$ for all $\omega^* = (\cdot, Bad)$. His subjective state becomes the objective state space Ω^*.

2.2 Decision-Theoretic Foundations

This section models unforeseen contingencies from the decision-theoretic viewpoint. The decision theoretic approach starts from the agent's preference or choice behavior, which is more relevant for economists.

2.2.1 (Bounded) Rationality

Rationality

Before introducing unforeseen contingencies by decision-theoretic approach, we have a review of rational choice theory. A comprehensive discussion is in Rubinstein (1998). We start by defining rationality. We can define rationality either procedurally or behaviorally.

- *Procedural definition*

Definition 2.1 *A **preference** is a binary relation $\succeq$ defined on a set of alternatives X.*

$x \succeq y$ is interpreted as the agent weakly prefers x to y, or x is at least as good as y.

We call $\succeq$ *complete* if, $\forall x, y \in X$, either $x \succeq y$ or $y \succeq x$. In words, the agent is able to compare any two alternatives.

We call $\succeq$ *transitive* if, $\forall x, y, z \in X$, $x \succeq y$ and $y \succeq z$ imply $x \succeq z$. Transitivity is the core feature of rationality. It rules out the cyclic preference. Otherwise, the agent is a "money-pump". Suppose $x \succeq y$, $y \succeq z$ and $z \succeq x$. We have all the alternatives x, y and z. Initially we give the agent z for free. Now we can exchange y for z by asking for some money. Then we can also exchange x for y by asking for some money. Then we can continue to exchange z for x. By doing it repeatedly, we can earn money from the agent until the agent's pocket is empty.

Definition 2.2 *$\succeq$ is **rational** if and only if $\succeq$ is a complete and transitive.*

For simplicity, assume that X is finite. $\forall A \subseteq X$ and $A \neq \emptyset$, a rational agent chooses the $\succeq$-maximal element in A.[2]

However, we cannot observe the preference directly. Hence, rationality of the preference has no empirical content and thus is not testable. But we can observe agents' behaviors and link the some regularity of the behaviors to the rationality of the underlying preference.

- *Behavioral definition*

Definition 2.3 *A choice function is $c(\cdot) : 2^X \mapsto X$ with $c(A) \in A$ for all $A \subseteq X$.*[3]

There are two types of consistency of the choice behavior.

Definition 2.4 *$c(\cdot)$ satisfies **Weak Axiom of Revealed Preference (WARP)** if and only if, $\forall A, B \in 2^X$, $c(A) \neq c(B)$ and $c(B) \in A$ implies $c(A) \notin B$.*

The intuition for WARP is that the best alternative $c(B)$ in B is also available in A. But, in A, the agent chooses a different alternative $c(A)$, so $c(A)$ should be better than $c(B)$. Thus $c(A)$ should be out of B. Otherwise, in B, the agent would choose $c(A)$ but not $c(B)$.

[2] $\succeq$ is independent of A.

[3] For simplicity, we focus on function rather than correspondence.

Definition 2.5 *$c(\cdot)$ satisfies* ***Independence of Irrelevant Alternatives (IIA)*** *if and only if, $\forall A,\ B \in 2^X$, $c(B) \in A \subset B$ implies $c(B) = c(A)$.*

IIA says, in the larger set B, the agent chooses $c(B)$. If we reduce B to A which contains also $c(B)$, then the agent should also chooses $c(B)$ in A.

The following theorem shows these two types of consistency of choice behavior are equivalent.

Theorem 2.5 *WARP is equivalent to IIA.*

Proof 2.5 $\Longrightarrow$: *Suppose $c(B) \in A \subset B$ and $c(B) \neq c(A)$. Then, by WARP, $c(A) \notin B$, a contradiction.*

$\Longleftarrow$: *Suppose $c(A) \neq c(B)$, $c(B) \in A$ and $c(A) \in B$. Thus $A \cap B \neq \emptyset$. Since $c(A) \in A \cap B$ and $c(B) \in A \cap B$, by IIA, $c(A) = c(B) = c(A \cap B)$, a contradiction.*

- *"as if" justification*

Now we construct a bridge between procedural rationality and behavioral rationality. Since we focus on function $c(\cdot)$ rather than correspondence, here we content ourselves with strict preference $\succ$, which is an asymmetric preference relation.[4]

Theorem 2.6 *$c(\cdot)$ satisfies IIA if and only if there exists a linear order $\succ$ on X s.t., $\forall A \subseteq X$, $c(A)$ is the $\succ$-maximal element in A.*

Proof 2.6 *It is straightforward to see the "if" part. Now we show the "only if" part. Suppose c satisfies IIA. We define $\succ$ as $a \succ b$ iff $c(a, b) = a$. Thus $\succ$ is complete. Suppose $a \succ b$ and $b \succ c$. Then $c(a, b, c) = a$. Otherwise IIA is violated. It implies $a \succ c$. Thus $\succ$ is transitive. Since c is a function, $\succ$ is asymmetric. Finally, by IIA, $c(A) = c(a, c(A))$ for all $a \in A$. Thus $c(A) \succ a$ for all $a \in A \setminus \{a\}$. Hence $\forall A \subseteq X$, $c(A)$ is the $\succ$-maximal element in A.*

"As if" justification: if an agent's choice behavior satisfies IIA then economists describe her behavior as if it is induced by a maximization procedure.

Example 2.1 *Simon's "satisficing" procedure.*

[4] $\succ$ is asymmetric if $x \succ y$ implies not $y \succ x$.

Let the primitives be $(X, S, >)$. *The procedure is as follows.* $\forall A \subseteq X$, *the agent goes over the elements in* A *according to the basic ordering* $>$ *defined on* X, *and picks the 1st element that belongs to* S. *If no element in* A *belongs to* S, *the agent selects the last element in* A *according to* $>$.

Obviously, this is not a maximization procedure, but the choice function $c(\cdot)$ *induced by this procedure satisfies IIA.*

To show it, let $A \subset B$ *and* $c(B) \in A$.

Case 1) The 1st element in $B \cap S$ *according to* $>$ *belongs to* A.

Since $A \subset B$, *this element is also the 1st element in* $A \cap S$ *according to* $>$, *so* $c(A) = c(B)$.

Case 2) $B \cap S$ *is empty.*

Hence the last element in B *according to* $>$ *is chosen; but this means that* $A \cap S$ *is also empty and the last element in* B *according to* $>$ *is also the last element in* A.

Since the choice behavior satisfies IIA, there exists a strict preference $\succ$ *that induces this choice behavior.*

Define the relation $\succ$ *where the following properties are satisfied:*

1. $\forall x_1, x_2 \in S$: $x_1 \succ x_2 \Leftrightarrow x_1 > x_2$.
2. $\forall x_1 \in S$, $\forall x_2 \notin S$: $x_1 \succ x_2$.
3. $\forall x_1, x_2 \notin S$: $x_1 \succ x_2 \Leftrightarrow x_1 < x_2$.

By definition, $\succ$ *is complete. It is also easy to show that* $\succ$ *is transitive. Thus* $\succ$ *is a rational preference relation.*

Now we show that $\succ$ *induces the same choice behavior as the satisficing procedure.*

It is equivalent to show that $\forall A \subset X$, $c(A)$ *is* $\succ$*-maximal element in* A.

a) Suppose $c(A) \in S$.

By definition of the satisficing procedure, $c(A)$ *is the 1st element in* $S \cap A$. *By property 1 in definition of* $\succ$, $\forall x \in A \cap S \setminus \{c(A)\}$, $c(A) \succ x$. *By property 2,* $\forall x \in A \setminus S$, $c(A) \succ x$. *Thus* $c(A)$ *is* $\succ$*-maximal element in* A.

b) Suppose $c(A) \notin S$, *then* $A \cap S = \emptyset$.

By definition of the satisficing procedure, $c(A)$ *is the last element in* A. *By property 3 in definition of* $\succ$, $\forall x \in A$, $c(A) \succ x$.

Thus $\forall A \subset X$, $c(A)$ *is* $\succ$*-maximal element in* A.

Since X *is finite, according to* $\succ$, *we can construct a utility function that represents this preference. Now we search for this utility function. Let* k *be the numbering of the element in* X *according to* $>$. *For example,* $\forall k \in S : u(k) = v^* + (|X| - k)C$

and $\forall k \notin S : u(k) = kC$ *with* $0 < C|X| < v^*$ *is a particular utility function whose maximization induces the same choice behavior as the satisficing procedure.* C *can be interpreted as the cost of searching.*

The example above shows that although some behavior is not based on any maximization procedure, but it could be rationalized by some rational preference. The idea of "as if" approach is introduced by Milton Friedman. He argues that theories should be judged only by their predictions but not assumptions.

However, if the human behavior systematically violates rationality, then "as if" approach breaks down.

Systematic Violations of Rationality

There are many examples of systematic violations of rationality. Here we only focus on the examples related to unawareness.[5]

There are two typical situations where the agent does not always make the same choice between two alternatives.

(i) *Time Inconsistent Preferences*

Example 2.2 *Dynamically inconsistent preferences: long-term planning (cold) vs. visceral urges (hot).*
For example, if "cold" is not aware of the existence of the "cold" self, she prefers to follow the study plan, but "hot" self ends up with the on line game without any restriction.

(ii) *Framing Effects*

Example 2.3 *Framing effects are widely observed and well-documented in psychology, and especially in cognitive psychology. The following optical illusion is a typical example.*

At first glance the left segment seems longer than the right one. However, the fact is that these two segments are equally long. The problem for us is that two tails are framed differently. In general, framing effect means that people's judgments of the same object will be different if this object is put into different context or described differently. It is not only confined in perception (truth judgment). It works also for value judgment. Consider the following two flight insurance contracts:

[5] A survey by Rabin (1998) and Rubinstein (1998) cover more examples.

$$\left(\textit{Death due to any reason}, \quad \$100,000\right)$$

$$\left(\begin{array}{cc} \textit{Death due to any terrorism}, & \$100,000 \\ \textit{any mechanical failure}, & \$100,000 \\ \textit{any other reasons}, & \$100,000 \end{array}\right).$$

The underlying mappings of these two contracts are the same. But psychological evidences show that people are willing to pay more for the second contract, since facing only the first one, most of them are not aware of "terrorism". Standard contract theory abstracts from framing effects. Because, in standard economics, we assume peoples' preferences are not affected by irrelevant features of the alternatives. Here framing contingencies play a role in people's preferences.

Rationalizations

The context-free rationalization of irrational behavior is considered by Kalai, Rubinstein and Spiegler (2002). They suggest *multiple rationales*.

Definition 2.6 *A* ***rationalization by multiple rationales (RMR) of the choice function*** *c is a K-tuple of strict preference relations $(\succ_k)_{k=1,\cdots,K}$ on X if, $\forall A$, $c(A)$ is $\succ_k$-maximal in A for some k.*

Let $r(c)$ be the minimal number of orderings among the RMR's of c.

Theorem 2.7 *$r(c) \leq |X| - 1$ for every choice function c.*

Proof 2.7 *Let $b \neq c(X)$. Let, $\forall a \neq b$, $\succ_a$ be a preference relation such that, $\forall x \in X \setminus \{a, b\}$, $a \succ_a b \succ_a x$.*

$\forall c(\cdot)$, $\forall A \subseteq X$, if $c(A) = a \neq b$, then $c(A)$ could be rationalized by $\succ_a$. If $c(A) = b$, then $A \subset X$. $c(A)$ could be rationalized by $\succ_x$ with $x \in X \setminus A$. Thus $c(\cdot)$ could be rationalized by rationales $\{\succ_a\}_{a \in X \setminus \{b\}}$. Moreover $|\{\succ_a\}_{a \in X \setminus \{b\}}| = |X| - 1$.

Example 2.4 *The (u, v) Procedure:*

The primitives are two functions $u : X \mapsto R$ and $v : X \mapsto R$ and a number v^*. The agent chooses an element that maximizes u as long as its v-value is at least v^*, and chooses an element that maximizes v otherwise.

Its interpretation is as follows: $u(\cdot)$ is the agent's subjective utility function when facing the recurrent similar decision situations without the need of awareness of some other *aspects of the situation.*[6] $v(\cdot)$ *is the relatively more objective utility function when more aspects of the situation are in the decision maker's mind.*

For example, let X *denote the set of natural number of cigarettes.* $u(x) = x$*,* $v(x) = -x$ *and* $v^* = 3$*.* $\forall A \subset X$ *with* $maxA \leq 3$*, the agent maximizes* $u(x)$*. He believes that the more the better. However, when* $maxA = 10000$*, a health aspect of the situation suddenly comes into his mind. He believes that if he consumes all cigarettes as he did as before, then he will die. Moreover, he is aware that smoking is a bad habit. Thus he updates his utility function from* u *to* v*.*

Furthermore, IIA is violated. For example, here $c(\{1, 2\}) = 2$*, and yet* $c(\{1, 2, 3, 4\}) = 1$*.*

2.2.2 Decision under Unforeseen Contingencies

Dekel *et. al.* (1998b) survey the models of unforeseen contingencies by decision-theoretic approach. Here we do not explore the agent's true belief, but investigate his preference, and then conclude that the agent behaves *as if* he holds some belief, probably with unforeseen contingencies.

The first relevant work is on Savage (1954)'s invention of *subjective probability*.

The state space is Ω. X is the set of consequences. The *act* is a function from Ω to X. The agent's preference is defined on the set of acts F. Savage shows that if the preference $\succeq$ satisfies certain properties, then there exists a function $u : X \to \mathbb{R}$ and a probability measure Q such that $f \succeq g$ iff $\sum_{\omega\in\Omega} Q(\omega)u(f(\omega)) \geq \sum_{\omega\in\Omega} Q(\omega)u(g(\omega))$ for all f, $g \in F$.

Thus from the agent's behavior, we derive his subjective probabilities. However, it is implicitly assumed that the agent understands the complete state space which rules out unforeseen contingencies. Now we introduce the notion of small worlds. The idea is that the subjective state space from the agent's view is a partition of the objective state space. Some particular state in the event in the partition is either irrelevant or unforeseen by the agent.

Consider the following example. Let $\Omega = \{(rain, revolution), (rain, no\ revolution), (no\ rain, revolution), (no\ rain, no\ revolution)\}$. There are two dimensions of uncertainty. One is whether it rains tomorrow. The other is whether there is a revolution in Haiti. If the latter dimension is unforeseen by the agent, then his subjective state space is $S = \{(rain), (no\ rain)\}$.

[6] The term "subjective" means subjectivity of belief but not subjectivity of preference. Belief-subjective utility could be wrong, since the individual "believes" the utility is of some form which is a hypotheses in his mind. On the other hand, we could not say preference-subjective utility is wrong, because the utility is the individual's true "feeling" which represents his personal value judgment.

The agent has three alternatives of actions: buy an umbrella, save the money and invest it in Haiti. Let the set of consequences X be the set of monetary values $\mathbb{R}$.

The three objective acts are following:

State	Consequence if Buy	Consequence if Save	Consequence if Invest
$(rain, revolution)$	5	0	-100
$(rain, no\ revolution)$	5	0	10
$(no\ rain, revolution)$	6	8	-100
$(no\ rain, no\ revolution)$	6	8	10

However, from the agent's perspective, three acts are following:

State	Perceived Consequence if Buy	Perceived Consequence if Save	Perceived Consequence if Invest
$(rain)$	5	0	x
$(no\ rain)$	6	8	x'

The problem is that the act of investing is not measurable with respect to his awareness. Thus x and x' are hard to determine.

Modica *et al.* (1998) provide a natural way to determine them. Since the dimension of revolution is unforeseen by the agent, he just implicitly assigns a default value of no revolution to this dimension. Thus $x = x' = 10$. However, this modeling method rules out that the agent is aware that there might be some unforeseen contingencies.

Ghirardato (2001) suggests that act is a correspondence that maps from S to $2^X \backslash \emptyset$. $f(s)$ is the set of consequences possible in s.

Thus the agent view the acts as following:

State	Perceived Consequence if Buy	Perceived Consequence if Save	Perceived Consequence if Invest
$(rain)$	$\{5\}$	$\{0\}$	$\{-100, 10\}$
$(no\ rain)$	$\{6\}$	$\{8\}$	$\{-100, 10\}$

Ghirardato generalizes Savage's result that if the preference $\succeq$ satisfies certain properties, then there exists a function $u : 2^X \backslash \emptyset \to \mathbb{R}$ and a probability measure Q on S such that $f \succeq g$ iff $\sum_{s \in S} Q(s)u(f(s)) \geq \sum_{s \in S} Q(s)u(g(s))$ for all f, $g \in F$.

While this approach solves the immeasurability problem, it has one gap compared to Savage's model. In Savage (1954), states and consequences are objectively provided to the agent. So we observe the agent's preference over the objective acts and derive his subjective utility and subjective probability measure. However, in Ghirardato (2001), both states and acts are subjective. It has no empirical contents.

Skiadas (1997) extends the problem above to Savage. Skiadas regards consequences are not observable as well. Hence the agent has preference over (act, event) pairs. That is, we replace $f \succeq g$ by $(f, E) \succeq (g, E')$ for E, $E' \subseteq S$. Give some axioms, there is a utility function $u(f, E) = \sum_{s \in E} \frac{Q(s)}{Q(E)} u(f, \{s\})$. Therefore, $u(f, \{s\})$ is obtained. We derive the utility of the consequence $f(s)$ from the preferences as shown by the numbers in the following table.

State	Utility if Buy	Utility if Save	Utility if Invest
$(rain)$	a	b	c
$(no\ rain)$	d	e	f

But the solution is not complete, since the numbers in the table are only the average value within the coarse event in the partition S.

It is worth mentioning that Kreps (1979) gives an alternative approach by considering preferences for flexibility.

Recent development is in Ahn and Ergin (2007). It generalizes standard subjective expected utility theory through a non-additive set function $v(\cdot)$. In economics and psychology literature, framing effect is confined within the field of framing consequences, for example, prospect theory. But how state space is framed also plays an important role in decision making under uncertainty. Most importantly, AU (Awareness of Unawareness) which is not expressible in most epistemic models is possible in this model. Although some epistemic models have solved the AU problem, they have no direct implication on how to model economic models with AU. The following example provides a direct insight in economic models.

Example 2.5 *Suppose* $S = \{a, b, c\}$, $X = \{0, 1\}$. *Let* $\Pi = \Pi^*$. $v(\{a\}) = v(\{b\}) = v(\{c\}) = 1$, $v(\{a, b\}) = 1 + \alpha$, $v(\{b, c\}) = 1 + \alpha$, $v(\{a, c\}) = 2$.

Obviously, if $\alpha \geq 0$, *then* $v(\cdot)$ *is monotone.*

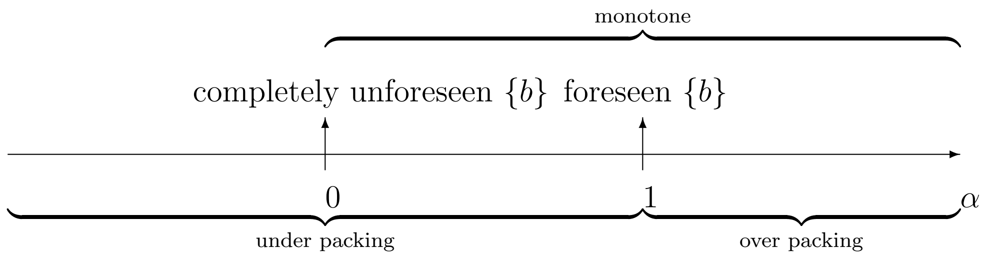
If $\alpha = 0$, *then* $\{b\}$ *is a completely unforeseen event. The reason is that* $\{b\}$ *is not null and* $\frac{v(\{a,b\})u(p)+v(\{c\})u(q)}{v(\{a,b\})+v(\{c\})} = \frac{u(p)+u(q)}{2} = \frac{v(\{a\})u(p)+v(\{b,c\})u(q)}{v(\{a\})+v(\{b,c\})}$ *for all* $p, q \in \Delta X$.

If $\alpha = 1$, *then* $\{b\}$ *and all other events are foreseen (by Proposition 6(i) in Ahn and Ergin (2007)). The model is degenerate to standard expected utility theory. On the other hand, if* $\alpha \neq 1$, *the family of all foreseen events* $\mathcal{A} = \emptyset$

Most interestingly, if $\alpha < 1$, *then* $v(\cdot)$ *is sub additive. If* $\alpha > 1$, *then* $v(\cdot)$ *is super additive. Moreover, in either case,* $\{b\}$ *is partially unforeseen.*

However, their model also involves some drawbacks.

Consider the following event list in Ahn and Ergin (2007).



$$\begin{pmatrix} \$500 & \text{Surgery} \\ \$100 & \text{Prenatal care} \end{pmatrix} = \begin{pmatrix} \$500 & E_1 \\ \$100 & E_2 \end{pmatrix}$$

First, $E_1 \cap E_2 \neq \emptyset$. Most contracts like this are incomplete, since some clauses are contradictory as shown in Table 2.1.

Table 2.1:

	Surgery	No surgery
Prenatal care	**???**	\$100
No prenatal care	\$500	\$0

Thus there is no partition for the preference to be partition-dependent. What is required is multi-dimensional model.

Second, assumption (f, π_1) is preferred to (g, π_2) if and only if $(f, \pi_1 \vee \pi_2)$ is preferred to $(g, \pi_1 \vee \pi_2)$ is strong. Consider the following example:

Example 2.6 Home insurances:

Based on example 2.5, we consider a situation an insurer proposes contracts to an insuree who has a partition-dependent expected utility function with $v(\cdot)$. The interpretation of the state space is: a is the the state of fire, b is the state of lightning and c is the state of no calamity.

Suppose the insurer proposes a contract f firstly.

$$f = \begin{pmatrix} x_1, & \{a, b\} \\ x_2, & \{c\} \end{pmatrix} = \begin{pmatrix} x_1, & \textit{calamity} \\ x_2, & \textit{no calamity} \end{pmatrix}$$

where $\pi(f) = \{\{a, b\}, \{c\}\}$.

Then the insuree contemplates proposing g or g'.

$$g = \begin{pmatrix} y_1, & \{a\} \\ y_2, & \{b, c\} \end{pmatrix} = \begin{pmatrix} y_1, & \textit{fire} \\ y_2, & \textit{no fire} \end{pmatrix}$$

where $\pi(g) = \{\{a\}, \{b, c\}\}$.

$$g' = \begin{pmatrix} y_1, & \{a\} \\ y_2, & \{b\} \\ y_2, & \{c\} \end{pmatrix} = \begin{pmatrix} y_1, & \textit{fire} \\ y_2, & \textit{lightning} \\ y_2, & \textit{no calamity} \end{pmatrix}$$

$\pi(g') = \{\{a\}, \{b\}, \{c\}\}$.

By the assumption, there is no difference between proposing g *and* g'*. since* $\pi(f) \vee \pi(g) = \pi(f) \vee \pi(g') = \{\{a\}, \{b\}, \{c\}\}$*, proposing either contracts leads a full awareness of the insuree for all* α*. But it is not so plausible that the insuree could be aware of lightning if* g *follows* f*. The reason is that the model does not distinguish directly mentioning lightning and indirectly stating calamity but not fire.*

Chapter 3

Unforeseen Contingencies and Incomplete Contracts

Why are contracts so incomplete in reality? It is simply because the benefit of incompleteness exceeds the cost of it for the drafting parties. Although it cannot be viewed as the full answer, it works as the way we analyze the problem.

Before providing the answer, it is necessary to define contractual incompleteness in the first place. In contract theory, contract is modeled in reduced form as a mapping, which maps from contingencies (or events) to the actions (obligations) of contracting parties. While there are many definitions for incomplete contracts in contract theory, we focus on the most plausible one. Incompleteness comes in two forms. First, an incomplete contract arises when the domain of the mapping is not a partition of the state space. Put it differently, there are gaps or contradictions. Second, the obligations specified in the contract do spell out all the pay-off relevant actions of the parties.

It is worth mentioning the so-called incomplete contracting paradigm initiated by Grossman and Hart (1986), Hart and Moore (1990). They assume that some contingency and ex-ante action (say, investment) are unforeseen or foreseen but at least not describable, which induces a hold-up problem ex post. By restricting the choice of contracts to only simple contracts, they therefore study the optimal allocation of asset ownership.[1] Aghion and Bolton (1992) apply this approach into financial contracting problems to determine the control rights. Later, Aghion and Tirole (1997) assume that ex post actions are not describable as well and endogenize the optimal authority via the complete contract paradigm.

Although the classic literature derives interesting economic implications, it is silent on

[1]Maskin and Tirole (1999) show that indescribability of two contingencies are irrelevant if these two contingencies can neither provide insurance for the parties ex post nor promote incentives ex ante via the implementation theory.

why those contingencies and actions are not describable. We now survey shortly some reasons for incomplete contracts:

1. Explicit Writing Costs

It is clear that there are benefits for writing complete contracts by the common wisdom. The reason why there are so many incomplete contracts around us is that some costs of completing contracts do exist for the drafting party.

Dye (1985) first introduces an exogenous writing cost per contingency to endogenize incomplete contracts. Anderlini and Felli (1999) follow this approach in a simple bilateral risk-sharing model.[2] The optimal sharing rule is incomplete in the sense that it does not prescribe the best allocations for some contingencies. Battigalli and Maggi (2002) separate writing costs into two parts: costs of describing the events and costs of describing the parties' actions. They predict two kinds of incompleteness: discretion, meaning that the actions of parties are not specified in details; and rigidity, meaning that the obligations of parties are not contingent on the detailed contingencies.

Shavell (2006) shows that when the court plays a benevolent role in interpreting contracts, contracts turn to be more incomplete, i.e., contracts include more gaps and fairly general terms, and the court should not always enforce what parties write in contracts. Heller and Spiegler (2007) endogenize a special kind of contractual incompleteness: contradiction, under the precedent system.

However, in all these models, given the writing cost per term, the drafting party optimally writes a contract. It is artificial in the sense that the writing cost here seems only the physical ink and paper. Thus it is more interesting to endogenize the explicit writing costs.

2. Thinking Costs

In contrast to the ink cost, a more intuitive way to model the cost of writing extra clauses to complete a contract is the cognitive or thinking cost. There are costs of searching for the contingencies.

In Bolton and Faure-Grimaud (2007), thinking ahead to write a complete plan involves delays of current decisions. So the writing cost of extra clauses is the opportunity of forgoing the current decisions. Interestingly, they show that the impatient party may give the control right to the more patient party.

Tirole (2008) uses a well-behaved cost function to model the thinking cost. The cost of thinking bijectively maps to the probability of finding the unforeseen contingency, which therefore becomes describable after being found out. It is also related to AU, which is an interesting issue in unawareness literature. We will discuss this model formally later.

[2]Besides ex ante writing costs, they also introduce ex post implementing costs.

3. Signaling

In the environment with asymmetric information between contracting parties, an incomplete contract can be a particular signaling device. The writing cost for the good drafting party here is endogenized by the cost of pooling with the bad party.

In Aghion and Bolton (1987), a good supplier specifies no penalty for breach of contract by the buyer who switches to a different supplier so as to signal that entry of potential competitors is unlikely. In Spier (1992), a good football player does not specify an injury insurance clause in her contract in order to signal an accident is unlikely. In Hermalin (2002), a good worker forgoes the penalty for breach by the employer so as to signal he is not afraid of going back to the labor market again. In Chung and Fortnow (2007), the law writer writes a simple contract to signal his low awareness to the interpreter.

4. Flexibility

In the environment where some contingencies or actions are not describable ex ante, flexibility of some verifiable action, say price, is valuable ex post.

In Hart and Moore (2008), if the contracting parties are generous and therefore will not inefficiently retaliate against each other ex post, they do not have to specify the trading price ex ante as the buyer's valuation and the seller's cost are both uncertain ex ante. Forgoing the price promotes a higher trading opportunity ex post. The writing cost of the price in the contract is the potential gain from trade due to the rigidity of the price.

Martimort and Piccolo (2008) study the contracting relationships between the suppliers and retailers. Since there are some indescribable non-market actions of the retailers, the supplier may forgo the retail prices in the contract to promote the efficient downstream competition ex post. The writing cost of the price in the contract is the opportunity cost of the horizontal externality between the retailers less the agency cost.

5. Strategic Shroud for Unaware Parties

Recent approaches endogenize incompleteness of contracts from psychology and economics. When one party is fully rational and the other party is boundedly rational, the contractual incompleteness can be as the result of strategic shrouding by the rational party.

Dekel *et. al.* (1998b) distinguish the difference between unforeseen contingencies and the events the agent has in mind but assigns zero probability. The agent's probability judgment of the former event is changed if an uninformative statement, say 'the event might happen or not', is announced to the agent. But the agent's judgment of the later is immune to such statement.

Besides their argument, Li (2008) argues that zero probability cannot capture the problem of unawareness of alternatives, the choice of which is endogenous.

In Gabaix and Laibson (2006), firms may shroud the add-on to exploit the unaware consumers ex post. In Filiz (2006), an insurer may shroud some contingencies to an unaware insuree. Hence the writing cost is endogenized by the drafting party's opportunity cost of exploiting the non-drafting party. It suggests to use mandatory contract terms ex ante, and therefore void such contracts as a policy response as in Korobkin (2003). However, von Thadden and Zhao (2007) show that the strategic shroud of some actions of the non-drafting party increases the social welfare, as it utilizes the non-drafting party's unawareness of some bad but non-verifiable actions. In this paper, the writing cost for the drafting party is endogenized by the cost of adding incentive constraints.

6. Vagueness

Vague terms is another form of contractual incompleteness.[3] Although vagueness is very prevalent in our natural language, sometimes real-life contracting parties purposefully write vague clauses, say "taking appropriate care" or "reporting at regular periods". (See Lipman, 2003) However, this is an unexplored field. The future research should provide a rationale for vagueness in contracts.

3.1 Completely Unforeseen Contingencies

3.1.1 Information Disclosure and Consumer Awareness

Adverse product effects are a serious economic problem. As a result of information disadvantages, consumers may be unaware of some low-quality aspects of products, for example, harmful radiation from computer monitors or cell-phones, health risks due to nanoparticles or artificial sweeteners in food, and side effects of medicines. A profit-seeking firm may use many ingredients in different degrees for the production of products. Such ingredients are supposed to improve the performance of the product or reduce the cost of production. However, these substances may have adverse effects on consumer well-being. Not only may such adverse effects be uncertain and of unknown degree but the consumer may initially be unaware that, by consuming, they expose them to such a risk. More examples are asbestos, nicotine, transgenic fats, and flavor enhancers, whose health risks are or were largely unknown by the consumers. The recent debate on genetically modified agricultural products has a similar flavor: Firms use products with certain genetic modifications; consumers are imperfectly informed about the degree of such modifications and whether such modifications are harmful.

In this paper, we consider two classes of information problems on the consumer side: uncertainty and unawareness. To model uncertainty on the consumer side, we develop

[3] A word is precise if it describes a well-defined set of objects, as a word is vague if it is not precise. For example, the words "appropriate" and "probably" are vague.

a simple fully Bayesian model. To model unawareness, we develop a simple model in which consumers suffer from a biased prior. The main contribution of the paper is that we provide a simple model that allows us to highlight the conceptual differences between consumer unawareness and consumer uncertainty.

Formally, with some probability, some characteristic of a product by a firm creates health problems for consumers. Otherwise, this characteristic does not affect the well-being of consumers. In the "uncertainty" model, consumers are aware of the substance but uncertain of the level of the substance and whether the substance is harmful. In the unawareness model consumers are not aware of the existence of the substance at all. The monopolist firm knows whether the substance is harmful or not and the level of it. He then decides about his disclosure policy: He may fully, partially, or not at all disclose information through advertising.

The main results are as follows. First, social welfare may be higher with unaware consumers than that with aware consumers. Intuitively, a monopolist always sets a price higher than the social optimal level. This leads to too little consumption in the market with complete information. Hiding information, however, leads to too much consumption. The distortions created by monopoly and hiding information go in the opposite directions. Hence, in the presence of monopoly power, hiding information is not necessarily detrimental to welfare. However, the conclusion with respect to consumer surplus is unambiguous: Consumers are always better off if they are aware.

Second, from a policy perspective, mandatory information disclosure makes unaware consumers better off. However, *consumers are potentially worse off if full instead of partial disclosure is mandated.* The reason is that imposing a mandatory full disclosure rule may lead to non-participation of the monopolist.

Related Literature: In information disclosure problems, uncertainty problems have the feature that consumers are uncertain in the sense that they know the distribution of the relevant unknown attribute, although they do not know the exact value the attribute takes. The underlying adverse selection problem can be solved through voluntary information disclosure by the firm. It is well-known that, if such disclosure is costless, full unravelling results and the adverse selection problem is fully solved. See, e.g., Grossman and Hart (1980), Grossman (1981), Milgrom (1981), Milgrom and Roberts (1986), and the generalized model by Okuno-Fujiwara, Postlewaite, and Suzumura (1990). However, as also holds in our setting, if disclosure is costly no or only partial unravelling will occur (see, e.g., Shavell, 1994). After disclosure, consumers update their beliefs in a Bayesian fashion upon observing firms' disclosure actions. With respect to the contracting literature on information disclosure, we refer to the overview provided in chapter 5 in Bolton and Dewatripont (2005).

The law and economics literature has used the above approach to address consumer protection issues.[4] In the legal literature, Korobkin (2003) recommends ex ante intervention by legislatures; this corresponds to mandatory information disclosure rule. However, this ex ante mechanism sometimes inefficiently excludes firms from the market as information disclosure is costly as we show in this paper. Polinsky and Shavell (2006) compare mandatory to voluntary disclosure rules in a setting in which both the firms decide whether to acquire information. They show that firms may have less incentive to acquire information under mandatory disclosure. We note that if the legislator can require the seller to disclose all the possibly harmful substances, it matters how the seller discloses the eye-opening information. If the seller puts the information only in fine print, the seller's action constitutes mis-selling if the information does not reach the consumers.

Other work has considered ex post judicial mechanisms. In the economics literature, Daughety and Reinganum (1995) and Daughety and Reinganum (2008) examine the firm's behavior when the firm is liable to make a payment in the event of harm. However, in our context, harm is often not contractible. Thus, this judicial mechanism has limited applicability. It is worth mentioning that some legal scholars suggest another ex-post judicial mechanism (see Korobkin, 2003, and Becher, 2008): By using the unconsionability doctrine to interpret contracts, contracts with unconscionable terms (which, thus, put one party at the mercy of the other) are not enforced. Unfortunately, this mechanism appears to be of little help in our context because its implementation is difficult in the presence of adverse effects.

According to the second class of informational problems, unaware consumers do not know the attribute and do not know that they do not know it and so on so forth. To analyze this class of information problems, one has to give up common knowledge of the game (and rationality), and assume a non-common prior between firm and consumers. Epistemic foundations are provided by Board and Chung (2006), Galanis (2007), Heifetz, Meier and Schipper (2006), and Li (2009) in the unawareness literature. From a normative viewpoint, the consumers' prior is biased, unless they are made aware. This non-common prior approach has been used in a number of recent behavioral-IO models—see, e.g., present-biased consumers (Della Vigna and Malmendier, 2004) and the extension to diversely naive consumers (Eliaz and Spiegler, 2006), consumers who are unaware of some options (Eliaz and Spiegler, 2008), consumers who are unaware of some add-ons (Gabaix and Laibson, 2006), analogy-based-reasoning consumers (Mullainathan et al., 2008), limited-recall consumers (Shapiro, 2006), consumers who are susceptible to the law of small numbers (Spiegler, 2006)).[5] Our paper adds to this literature by taking a

[4] See, e.g., Shavell (2004) for extensive discussions of the law and economics literature on this issue. A different remedy with respect to adverse effects is to define minimum quality standards that refer to product safety or product quality (see, e.g. Leland, 1979, and Shapiro, 1983).

[5] In psychology, the related concept of awareness is availability (See Kahneman and Tversky, 1973). For an alternative Bayesian approach of modeling contracting with unawareness, see Tirole (2009).

closer look at information disclosure rules, highlighting the difference between a market inhabited by consumers that lack information but do not have biased beliefs and one in which consumers do have biased beliefs.

The rest of the paper is organized as follows. Section 3.1.1 presents the model with aware consumers and unaware consumers, respectively. Section 3.1.1 examines the welfare consequence and discusses the information disclosure policies. The last section concludes.

Information Disclosure with Aware and Unaware Consumers

The Model We present an adverse selection model in which a monopolist sells a single product to a unit mass of consumers. Consumers are aware of the existence of the product and know the utility from its intended use. However, the firm's product contains a potentially harmful substance (e.g., asbestos, nicotine, artificial sweeteners, genetically modified products etc.). The monopolist incurs constant marginal costs of production that are normalized to zero. He sets his price (or, equivalently, quantity) and his information disclosure policy, as will be specified below. We assume that the firm knows the exact quality of the product (i.e., whether or not the substance is harmful and which amount of it is used). Thus we rule out the problem of quality test (see, e.g., Matthews and Postlewaite, 1985).

We introduce the possibility of unawareness about adverse effects into a linear-quadratic representative-consumer model. In the context of information disclosure policies, this model has been used by Daughety and Reinganum (2005).[6] We refer to the model with aware consumers if consumers are aware of the potentially adverse effect; however, absent information disclosure, they lack information about whether such adverse effects are present and about the magnitude of these effects. We refer to the model with unaware consumers if consumers are not aware that there are potentially adverse effects, unless such information is disclosed. In effect, they have a biased prior.

We aim at developing a simple framework to analyze the difference of market environments with aware *vs.* unaware consumers. To do so, we need some notation: We denote

- θ as the amount of the substance, uniformly drawn from $[0, 1]$;
- I as an indicator which takes value $I = 1$ if the substance is harmful and $I = 0$ otherwise;

Zhao (2009) extends Tirole (2009) to a model with asymmetric awareness between a seller and a buyer and focuses on the transaction cost of pre-contractual cognition of the buyer. By contrast, in this paper, unaware consumers are biased in the sense they are naive; thus there is no cognition.

[6] We have checked that our results still hold in a heterogeneous consumer model with unit demand and a uniform distribution of the willingness-to-pay.

- p as the per-unit price of the product and q as the quantity sold by the monopolist.
- x as the probability that the substance is harmful;
- $a \equiv \alpha - I\theta$ as the true quality measure (net of any adverse effects) of the product, where α is a parameter that shifts the willingness-to-pay function and reflects the consumer's preference ignoring possible adverse effects;
- $\widetilde{a}$ as the consumers' expected quality.

For simplicity, we assume that θ and I are independent. We postulate that the utility function of the representative consumer takes the standard linear-quadratic form

$$U = (\alpha - I\theta)q - \frac{1}{2}q^2 - pq.$$

The firm may disclose information through advertising at a fixed cost $c > 0$. Advertising is, thus, by definition, truthful.[7]

The timing of the game played by monopolist and consumers evolves as follows:

1. Nature chooses θ and I.
2. The monopolist observes θ and I—this is his *private* information. He then chooses if it partially or fully discloses information through advertising and sets its price p. If the firm advertises, it chooses to disclose θ and/or I.
3. Consumers observe the price and, if applicable, the advertisement and then make their purchasing decision.

Notice that the consumer's decisions only depend on the consumer's expected quality. It is straightforward to obtain the following lemma:

Lemma 3.1 *Independent of whether the representative consumer is aware or not, we distinguish the following two cases:*

[7] This can be motivated by measures taken against misleading or false advertising. Such advertising about product characteristics is thus within the domain of informative advertising. However, since unaware consumers have biased beliefs advertising changes consumer preferences for the product at the moment of purchase—this is a feature of persuasive advertising. In contrast to work on persuasive advertising, in our setting advertising "corrects" consumer preferences—i.e., ex post preferences are the true preferences. For a monopoly model of persuasive advertising that allows for distorted preferences ex ante or ex post, see Dixit and Normann (1978); for a survey on the economics of advertising see Bagwell (2007).

1. *If $\widetilde{a} > 0$—i.e., the representative consumer's expected quality is greater than zero—the consumer buys a strictly positive quantity of the product. The equilibrium price is $p = \frac{\widetilde{a}}{2}$ and quantity is $q = \frac{\widetilde{a}}{2}$; the gross profit of the firm is $\frac{\widetilde{a}^2}{4}$. In equilibrium, the representative consumer's utility level is*

$$CS^{\widetilde{a},a} = \frac{a}{2}\widetilde{a} - \frac{3}{8}\widetilde{a}^2,$$

and the total surplus is

$$TS^{\widetilde{a},a} = \frac{a}{2}\widetilde{a} - \frac{1}{8}\widetilde{a}^2.$$

2. *If $\widetilde{a} \leq 0$, the firm does not produce.*

Notice that in the case where consumers buy the product, adopting a consumer welfare standard, it is optimal to have $\widetilde{a} = 2a/3$. The consumers should have a downward bias in this belief about product quality, $\widetilde{a} < a$, due to asymmetric information between the firm and consumers. However, adopting a total welfare standard, it would be optimal for the consumer to have biased beliefs $\widetilde{a} = 2a$. The reason is that the consumer's upward bias in the belief about quality, $\widetilde{a} < a$, counteracts the social underproduction in monopoly that would result under unbiased beliefs.

Aware Consumers As a benchmark model let us first analyze the model under the assumption that consumers are aware of the substance but *uncertain* of the level of θ and its presence in the product I.

Denote $\widetilde{a}_N$ the consumers' expected quality level in the absence of advertising.

By the unravelling argument (see, e.g., Milgrom, 1981, and Milgrom and Roberts, 1986), if the firm advertises, the firm will disclose both I and θ; note that there would be full unravelling if $c = 0$. Thus, the firm with quality a will advertise if and only if

$$\frac{a^2}{4} - c \geq \max\{0, \frac{\widetilde{a}_N^2}{4}\}.$$

There are two cases to consider, depending on whether $\widetilde{a}_N$ is positive or not. Denote $\widehat{a}$ the cutoff value of the quality at which the firm is indifferent between advertising and no advertising.

Case 1: $\widetilde{a}_N \geq 0$. Then the firm with quality a advertises if

$$a \geq \widehat{a} = \sqrt{4c + \widetilde{a}_N^2}. \tag{3.1}$$

To make things interesting, we assume that α is not too small:

Assumption 3.1 $\alpha > \frac{16c+1}{4}$

The assumption rules out the trivial case in which the cost of information disclosure c is too large such that the firm will never advertise.[8]

By Bayesian rule, the consumers' conditional expectation is (using the uniformity assumption)

$$\widetilde{a}_N = E[a|a \leq \widehat{a}] = \frac{\widehat{a} + \alpha - 1}{2}. \tag{3.2}$$

If c is not too large, the firm advertises with positive probability—i.e., with realizations of a above the critical value $\widehat{a}$ in the solution. Combining expressions (3.1) and (3.2), we obtain

$$\widetilde{a}_N = \frac{2}{3}\alpha + \frac{1}{3}\sqrt{12c - 2\alpha + \alpha^2 + 1} - \frac{2}{3} \tag{3.3}$$

In case 1, the consumer consumes a positive amount of the good at the profit-maximizing price. Equation (3.3) implies that $\widetilde{a}_N \geq 0$ if and only if

$$1 - 2\sqrt{c} \leq \alpha, \tag{3.4}$$

i.e., the cost of advertising is not too high compared to the highest quality of the product.

We now have that

$$\widehat{a} = \frac{1}{3}\alpha + \frac{2}{3}\sqrt{12c - 2\alpha + \alpha^2 + 1} - \frac{1}{3}.$$

Notice that $\widehat{a}$ is increasing in c. The advertising cost reduces the probability of information disclosure ex ante. Clearly, if c is equal to zero, there is full information disclosure due to unravelling. Note that $\widehat{a}$ does not depend on x since the unraveling logic implies that the consumer knows for sure the substance is harmful if there is no advertising.

Lemma 2 implies the following results:

If $I = 0$, the firm discloses I and θ and consumers learn that $a = \alpha$. The firm's net profit is $\frac{\alpha^2}{4} - c$, and the consumer's net utility level is $\frac{\alpha^2}{8}$.

If $I = 1$ and $a > \widehat{a}$, (i.e. θ sufficiently small) the firm discloses I and θ and consumers learn a. The firm's profit is $\frac{a^2}{4} - c$, and the consumers' utility level is $\frac{a^2}{8}$.

[8]When c is not too large advertising by some firm types takes place if and only if $\alpha > \sqrt{4c + \widetilde{a}_N^2}$. For the firm with $a = \alpha$ to have an strict incentive to advertise, $\widetilde{a}_N = \alpha - \frac{1}{2}$. Hence, we must have $\alpha > \sqrt{4c + \left(\alpha - \frac{1}{2}\right)^2}$ which is equivalent to $\alpha > (16c + 1)/4$.

If $I = 1$ and $a < \widehat{a}$, the firm does not advertise. As follows from Lemma 1, the firm's profit is $\frac{\widetilde{a}_N^2}{4}$, and the consumers' utility level is $\frac{a\widetilde{a}_N}{2} - \frac{3\widetilde{a}_N^2}{8}$.

The expected consumer surplus is

$$CS_{A1} = x\left(\int_{\alpha-1}^{\widehat{a}} \left(\frac{a\widetilde{a}_N}{2} - \frac{3\widetilde{a}_N^2}{8}\right) da + \int_{\widehat{a}}^{\alpha} \frac{a^2}{8} da\right) + (1-x)\,\frac{\alpha^2}{8}. \tag{3.5}$$

The expected total surplus is

$$TS_{A1} = x\left(\int_{\alpha-1}^{\widehat{a}} \left(\frac{\widetilde{a}_N^2}{4} + \frac{a\widetilde{a}_N}{2} - \frac{3\widetilde{a}_N^2}{8}\right) da + \int_{\widehat{a}}^{\alpha} \left(\frac{a^2}{4} - c + \frac{a^2}{8}\right) da\right)$$
$$+ (1-x)\left(\frac{\alpha^2}{4} - c + \frac{\alpha^2}{8}\right). \tag{3.6}$$

Case 2: $\widetilde{a}_N < 0$. In this case, if the firm does not advertise, the consumers will buy zero quantity. The firm advertises if and only if its profit after advertising $\frac{a^2}{4} - c$ is positive, or equivalently, $a > \widehat{a}$ with $\widehat{a} = 2\sqrt{c}$. If there is a positive probability of advertising ex ante, it is necessary that $\widehat{a} < \alpha$. For this to be the case, we have to assume that $\alpha > 2\sqrt{c}$ which is implied by Assumption 3.1. The reason is that if the firm will disclose information if consumers' expected quality is positive, the firm with the same quality will also disclose information if consumers' expected quality is negative. The expected consumer surplus is

$$CS_{A2} = x\int_{\widehat{a}}^{\alpha} \frac{a^2}{8} da + (1-x)\,\frac{\alpha^2}{8}. \tag{3.7}$$

The expected total surplus in this case is

$$TS_{A2} = x\int_{\widehat{a}}^{\alpha} \left(\frac{a^2}{4} - c + \frac{a^2}{8}\right) da + (1-x)\left(\frac{\alpha^2}{4} - c + \frac{\alpha^2}{8}\right). \tag{3.8}$$

Combining cases 1 and 2, we define the consumer surplus measure as

$$CS_A \equiv \begin{cases} CS_{A1} & \text{if } \widetilde{a}_N \geq 0, \\ CS_{A2} & \text{if } \widetilde{a}_N < 0. \end{cases}$$

We return to these surplus measures when comparing market environments in which consumers are aware to those in which they are unaware.

Unaware Consumers The analysis with unaware consumers is straightforward. Absent information disclosure, consumers are unaware of the potential adverse effect—i.e., consumers naively believe that $\widetilde{a} = \alpha$ if there is no advertisement about the substance. Therefore, no advertisement leads to the firm's maximal net profit $\frac{\alpha^2}{4}$. Hence, the firm does not make advertise in a market with unaware consumers. Thus in the second class of information problems there is zero advertising independent of the cost level. The firm does not have an incentive to solve the information problem consumers face because they are not aware of it.

If $I = 0$, the consumer's ex post utility level is $\frac{\alpha^2}{8}$. If $I = 1$, the consumers' ex post utility level is $\frac{a\alpha}{2} - \frac{3\alpha^2}{8}$, as follows from Lemma 1.

The expected consumer surplus is

$$CS_U = x \int_{\alpha-1}^{\alpha} \left(\frac{a\alpha}{2} - \frac{3\alpha^2}{8} \right) da + (1-x)\, \frac{\alpha^2}{8}. \tag{3.9}$$

The expected total surplus is

$$TS_U = x \int_{\alpha-1}^{\alpha} \left(\frac{a\alpha}{2} - \frac{\alpha^2}{8} \right) da + (1-x) \left(\frac{\alpha^2}{4} + \frac{\alpha^2}{8} \right). \tag{3.10}$$

Surplus Comparison and Mandatory Disclosure Rules

Surplus Comparison In this section, we obtain an ambiguous result about the impact of consumer welfare on total surplus. However, consumers are always better off if they are aware.

Proposition 3.1 *Welfare may increase or decrease if all consumers become aware—i.e., $TS_U - TS_A$ is of ambiguous sign— while consumers are better off if they are aware—i.e., $CS_U < CS_A$.*

Proof 3.1 *See Appendix A.1.1.*

The intuition that social welfare may be higher with unaware consumers than that with aware consumers is simple. We note that a monopoly seller always sets a price higher than the social optimal level. This leads to too little consumption in the market with complete information. Hiding information, however, leads to too much consumption because consumers are unaware of adverse effects that shifts their willingness-to-pay function downward. The distortions created by monopoly and hidden information go in opposite directions. Hence, in the presence of monopoly power, hiding information is not necessarily detrimental to welfare. Consequently, whether total surplus with aware consumers exceeds that with unaware consumers depends on the value the parameters take. If the bias of the consumer is small—i.e., x is small—or, relative to the scale of the harmful substance, the quality α is large, then the total surplus with unaware consumer is larger than with aware consumers. On the other hand, if the bias of the consumer is large and the quality is small relative to the scale of harmful substance, then the total surplus with aware consumer may be larger.

The conclusion with respect to consumer surplus is unambiguous: $CS_U < CS_A$. It does not depend on our specific assumption about the parameters. Aware consumers always obtain a higher surplus than unaware consumers: in comparison to aware consumers unaware consumers purchase too much and thereby have a lower surplus—this holds in equilibrium. To summarize, while unawareness of adverse effects counteract monopoly distortions from a total welfare perspective, consumers are always worse off if they are unaware.

Mandatory Disclosure Rules In this section we turn to consumer protection policies that may be introduced by the regulator or consumer protection authority. In particular, we explore the implications of two different information disclosure policies. We distinguish between *mandatory full information disclosure*, according to which the firm must reveal all the information that it has—i.e., I and θ. The resulting situation is one of full information. Alternatively, the firm may only be forced to reveal I (or, equivalently, the public authority performs its own analysis and reveals the realization of I to consumer.) With this *mandatory partial information disclosure* in place, consumers learn whether a substance if harmful but the firm is not required to reveal the amount of the substance that is contained in the product. Note that the model with mandatory partial information disclosure is formally equivalent to our previous model with aware consumers.[9] Thus, to evaluate the impact of information disclosure on consumers, we have to compare consumer surplus of the three models analyze above: the model with fully informed consumers, aware (but uninformed) consumers, and unaware consumers.

Assume first that there is a mandatory information disclosure rule such that the firm is required to disclose all its information. Mandatory information disclosure then leads to the full-information outcome. The firm's profit is $\frac{a^2}{4} - c$ for $a \geq \sqrt{4c}$ and zero otherwise, and the consumer's utility level is $\frac{a^2}{8}$.

Under condition (3.4), the firm will sell under mandatory disclosure of I and θ independent of its type a. Then the consumer surplus CS_M under mandatory information disclosure rule is always greater than CS_A. To prove this, notice that CS_A only depends on the threshold value $\widehat{a}$, and $CS_M = CS_A \mid_{\widehat{a}=0}$. Hence, we only need to prove CS_A is

[9] Recall that we call consumers aware if they are aware of the substance but uncertain of the value I and the level of θ. However, the regulator's disclosure of the harm I does not play any role for aware consumers because if $I = 0$, the firm will disclose this itself. Suppose $I = 1$, if the firm does not disclose θ, consumers know that the substance is harmful. The cutoff value is the same as before.

If consumers are unaware, disclosing I makes all consumers aware whether or not the substance is harmful. Hence, the regulator's disclosure of the harm is equivalent to the policy of making all consumers aware.

a decreasing function of $\widehat{a}$, as

$$
\begin{aligned}
\frac{\partial CS_A}{\partial \widehat{a}} &= x\left(\frac{\widehat{a}\widetilde{a}_N}{2} - \frac{3\widetilde{a}_N^2}{8} + \int_{\alpha-1}^{\widehat{a}}\left(\frac{a}{4} - \frac{3\widetilde{a}_N}{8}\right)da - \frac{\widehat{a}^2}{8}da\right) \\
&= -\frac{x\left(\widetilde{a}_N - (\alpha-1)\right)^2}{8} < 0.
\end{aligned}
$$

However, when condition (3.4) is violated, the firm does not sell if it is of sufficiently low quality and expected consumer surplus is

$$
\begin{aligned}
CS_M &= x\int_{\sqrt{4c}}^{\alpha}\frac{a^2}{8}da + (1-x)\frac{\alpha^2}{8} \\
&= \frac{(1-x)\,3\alpha^2 + x\alpha^3 - 8c^{\frac{3}{2}}x}{24}.
\end{aligned}
$$

The question is which disclosure policy is better. Here, we obtain the surprising result that more mandated information disclosure is not necessarily beneficial to consumers.

Proposition 3.2 *Mandatory information disclosure makes unaware consumers better off. However, consumers are potentially worse off if full instead of partial disclosure is mandated.*

Proof 3.2 *In case 1 (i.e., $\widetilde{a}_N \geq 0$),*

$$CS_M - CS_A = -\frac{xB_1}{648}$$

where

$$
\begin{aligned}
B_1 \equiv\ & 216c^{\frac{3}{2}} - (\alpha-1)\left(19(\alpha-1)^2 - 72c\right) \\
& - 6(\alpha-1)^2\sqrt{12c + (\alpha-1)^2} - 2\left(12c + (\alpha-1)^2\right)^{\frac{3}{2}}
\end{aligned}
$$

which can be positive or negative.

In case 2 (i.e., $\widetilde{a}_N < 0$),

$$CS_M - CS_A = -\frac{xB_2}{648}$$

where

$$B_2 \equiv 216c^{\frac{3}{2}} - (\alpha-1)\left(13(\alpha-1)^2 + 144c\right) - \left(14(\alpha-1)^2 + 96c\right)\sqrt{12c + (\alpha-1)^2}$$

which is always negative. We also recall that $CS_A > CS_U$. Thus consumers also benefit from information disclosure to the realization of I only. With respect to full disclosure we observe that $CS_M > CS_U$ only holds for a subset of the parameter space.

$$CS_M - CS_U = -\frac{xB_3}{24},$$

where

$$B_3 \equiv 8c^{\frac{3}{2}} - \alpha^3 - 6\alpha + 3\alpha^2.$$

The condition $B_3 < 0$ is equivalent to

$$c < (\frac{3}{4}\alpha - \frac{3}{8}\alpha^2 + \frac{1}{8}\alpha^3)^{\frac{2}{3}}.$$

We can show that, under Assumption 1, $B_3 < 0$.

Thus, in case 1, when $B_3 < 0$ and $B_1 > 0$, we have that $CS_U < CS_M < CS_A$. This means that mandatory information disclosure leads to a larger consumer surplus for ex ante unaware consumers. However, consumers gain if, instead of full disclosure, only partial disclosure is mandated. Overall the following two orderings are possible: (i) $CS_U < CS_M < CS_A$ and (ii) $CS_U < CS_A < CS_M$.

The reason of our surprising result that full disclosure can be worse than partial disclosure is that imposing a mandatory disclosure rule may lead to non-participation of the monopolist. If the monopolist was not allowed to quit the market (and, thus, his participation constraint would be ignored), we always would have $CS_A < CS_M$—this holds for the same reason as in the case that condition (3.4) holds.

Our result can be given a different interpretation: If $CS_U < CS_M < CS_A$, mandatory full information disclosure makes aware consumers worse off, while it makes unaware consumers better off. For the other two orderings the qualitative effect of mandatory full disclosure is the same for aware and unaware consumers alike.

Discussion and Conclusion

This paper presented a simple monopoly model to compare the effect of a potentially harmful substance in a market with aware in contrast to unaware consumers. We found that total surplus may be larger if consumers are unaware of the harmful substance. This makes them buy too much, which partly corrects for the underconsumption under monopoly. More importantly, we show that full mandatory disclosure may be harmful in the context of unaware consumers and that partial mandatory disclosure may be welfare-superior.

We motivated our analysis by referring to potentially harmful substances. More generally, our analysis applies to products which affect consumers' utilities although they may not be aware of this at the moment of purchase. In particular, it applies to complex products about which information is, in principle, available, but about which consumers may suffer from biased beliefs at the moment of purchase. For instance, consumers may be adversely surprised by the add-on costs of cartridges after buying a printer, or of using the telephone, or of watching in-room movies in a hotel. Public agencies are concerned about mis-selling: The Office of Communications in the United Kingdom found mis-selling in telecom services to be a growing problem.[10] Also, consumers in financial services are increasingly exposed to the mis-selling of complex financial products, such as endowment mortgages, private pensions, investment funds and insurance products. The Financial Services Authority noted as early as 2000 that one in eight consumers in the United Kingdom who had bought a financial product in the past five years later regretted her choice.[11] More recently, thousands of people in Hong Kong, Singapore and Taiwan took to the streets to protest and demand a refund of the money they lost from the financial products backed by failed Lehman Brothers in the financial crisis.[12]

Whether a certain action is to be considered mis-selling depends on consumer behavior. In a Bayesian world, consumers may lack information but they use correct beliefs given their information. Therefore, they cannot be systematically misled. This also means that non-disclosure and other attempts to hide unfavorable information, does not lead to systematically wrong purchase decision. By contrast, if consumers are unaware of certain product characteristics, the possibility of mis-selling arises. Here, information may be systematically suppressed by a firm. In this context, mandatory testing and disclosure rules are an important policy instrument to protect consumers. In case of non-compliance harsh punishments may be the only means to deter a firm from ignoring such consumer protection policies. Attempts to encourage information gathering by consumers have little relevance if consumers are completely naive in the sense that they are over-confident about their knowledge of the products and believe that nothing of the products will go wrong.

Our theory not only applies to how a product directly affects consumers but to the type of production processes that is used and the type on labor contracting within the firm and in vertical supply relationships. To be applicable, the utility that a consumer derives must depend on the use of inputs and contracts that the firm uses. This is the case if the utility function reflects ethical and environmental concerns. Examples are the disrespect of standards in labor contracts such as the use of child labor or forced labor (product examples: hand-woven carpets and textiles; concrete example: reports on sweat shops for products by NIKE) or the health and safety risks for workers

[10]See *Protecting citizen-consumers from mis-selling of fixed-line telecoms services,* Office of Communications, UK, 22 November 2004.

[11]See *Informed decisions? How consumers use Key Features: a synthesis of research on the use of product information at the point of sale,* Financial Services Authority, November 2000.

[12]See the article “Troubled Securities in Asia” in *The Economist*, November 20th 2008.

(product examples: mining products, textiles; concrete example: jeans dying in Turkey), disrespect of environmental standards (product examples: textiles, cleaning products,), disrespect of indigenous rights (product example: oil extraction for petrol), and animal experiments (product example: cosmetics). A concrete example are the war diamonds from the Congo; here consumers are concerned about the effect of upstream profits on the suffering of people, as a consequence of war that is financed through these profits. In this context NGOs play an important role in raising awareness; a firm's response consists in certifying the origin of inputs: De Beers certification efforts of the origin of its diamonds can be seen as a response to the consumer awareness that they may be buying a war diamond (which does not make a nice wedding gift). Under partial mandatory information disclosure, the government (or NGOs) makes consumers aware of the possible disrespect of certain standards. Such awareness campaigns make consumers aware of the relevance of a certain product characteristic that enters the consumer's utility function. It is up to the firms to certify that they follow certain business practices and comply with the standard. Such processes are often certified by third parties. In this sense, our paper shows a potential complementarity between mandatory information disclosure and private certification efforts: Partial public information disclosure may be necessary to make private certification efforts viable in market equilibrium.[13]

In this paper, we did not consider the situation of a mix of aware and unaware consumers. In such an extension it is interesting to study what happens if a larger share of consumers becomes aware of adverse effects—such a change in the composition of the population can come from public awareness campaigns that increase the share of aware consumers. The firm responds to a larger share of aware consumers by lowering its price. This benefits all consumers including those who remain unaware. Possibly an additional effect comes into play: if there is a sufficiently large share of aware consumers the firm advertises provided its product has a sufficiently small amount of the ingredient leading to an adverse effect.[14] In such a situation everybody becomes aware. Advertising leads to a further reduction in price and avoids the overconsumption that otherwise would prevail for unaware consumers.

[13]Whether private certification is fully revealing is a different issue. See Biglaiser (1993) and Lizzeri (1999) on this issue.

[14]A similar effect is also present in the work by Gabaix and Laibson (2006) who consider a competitive market in which some consumer are unaware of add-ons. They show that firms are more likely to disclose the add-on if the fraction of aware consumers in the population is higher.

3.1.2 Forward Induction and Games with Strategic Announcement

The game considered by Ozbay (2006) is as follows. There are two players: an announcer (he) and a decision maker, DM (she), indexed by P and A. Denote the objective probability space by $(\Omega, \Pr)$ where Ω is finite with $\Pr(\omega) > 0$ for all $\omega \in \Omega$. We assume the announcer knows $(\Omega, \Pr)$. But DM is only aware of the states in $\Omega_0 \subset \Omega$. Thus her subjective probability space is $(\Omega_0, Q_\emptyset)$ with $Q_\emptyset(\omega) = \Pr(\omega|\Omega_0)$ for all $\omega \in \Omega_0$. Here unawareness is modeled as incomplete state space, so the issue of awareness of unawareness is ruled out. The timing of their interaction is following:

• Stage 1: Nature moves.

Nature selects $\omega \in \Omega$. Both announcer and DM are uncertain that which state is realized. But the information set of the announcer is Ω while the information set of DM is Ω_0.

• Stage 2: Announcement of the announcer.

The announcer knows that DM is only aware of Ω_0. The strategy of the announcer is choosing the announcement $V \in 2^{\Omega\setminus\Omega_0}$ to make DM more aware. After DM's awareness is updated by V, DM's information set becomes $\Omega_0 \cup V$ on which a subjective probability Q_V is defined.[15] We assume that, for all $V \in 2^{\Omega\setminus\Omega_0}$ and $\omega \in \Omega_0 \cup V$, $Q_V(\omega) > 0$. This implies, once a new state is announced to DM, DM believes it is possible. Furthermore, we assume that, for all $V \in 2^{\Omega\setminus\Omega_0}$ and $\omega \in \Omega_0$, $Q_V(\omega|\Omega_0) = \Pr(\omega|\Omega_0)$. That is, the announcement does not alter the relative weights of the states in Ω_0.

• Stage 3: Action of the DM.

The strategy of DM is a decision function $d : 2^{\Omega\setminus\Omega_0} \longmapsto S$ where S is DM's action set.[16]

• Stage 4: Realization of the state of nature.

ω materializes. Their von Neuman-Morgenstein utility function are state-dependent $u_i : \Omega \times S \longrightarrow \mathbb{R}$ for $i = P, A$.

However, before this stage, they are both expected utility maximizers. The expected utility of the announcer is

$$U_P(a) \equiv \sum_{\omega\in\Omega} u_P(\omega, a) \Pr(\omega)$$

[15] Note that if $V = \emptyset$, $Q_V = Q_\emptyset$.

[16] Precisely speaking, d is not DM's strategy, because the DM cannot make a plan for the unforeseen contingencies. In fact, d represents the announcer's perceived response of DM.

and the expected utility of DM is

$$U_A(V, s|Q_V) \equiv \sum_{\omega \in \Omega_0 \cup V} u_A(\omega, s) Q_V(\omega).$$

Note that the expected utility of the announcer is an objective one while the expected utility of DM is subjective because of DM's subjective probability.

Definition 3.1 *An assessment* (V^*, d, Q) *is* ***rational*** *iff*
$V^* \in \arg\max_{V \in 2^{\Omega \setminus \Omega_0}} U_P(d(V))$ *and*
$d(V) \in \arg\max_{s \in S} U_A(V, s|Q_V)$ *for all* $V \in 2^{\Omega \setminus \Omega_0}$.

An *assessment* (V^*, d, Q) is rational iff the announcer chooses the optimal announcement given DM's decision function and DM chooses the optimal action given any announcement.

Definition 3.2 *An assessment* (V^*, d, Q) *is* ***justifiable*** *iff*
$\sum_{\omega \in \Omega_0 \cup V^*} u_P(\omega, d(V^*)) Q_{V^*}(\omega) \geq \sum_{\omega \in \Omega_0 \cup V^*} u_A(\omega, d(V)) Q_{V^*}(\omega)$ *for all* $V \subseteq V^*$.

Put it differently, DM makes a higher order reasoning. She believes that the announcer makes the optimal announcement. That is, DM believes that the announcer cannot improve his expected utility by announcing less.

Definition 3.3 *An assessment* (V, d, Q) *is* ***awareness equilibrium*** *iff it is rational and justifiable.*

Theorem 3.1 *Awareness equilibrium always exists.*

Proof 3.3 *The equilibrium is* $V^* = \emptyset$*, for all* V*,* $d(V)$ *is best action for the announcer among the actions that maximize DM's expected utility when* $V = \emptyset$ *and* Q_V *assigns a small probability to all* V *to guarantee that* $d(V)$ *is still the best action for DM. Since* Ω *and* S *are finite, the probability distribution exists. Thus* (V^*, d, Q) *is rational. Moreover, since* $V^* = \emptyset$*,* (V^*, d, Q) *is also justifiable.*

Ozbay (2006) shows there is always an awareness equilibrium where the announcer makes no announcement at all.

In addition, Ozbay (2006) also studies the equilibrium with reasoning refinement, which means when new contingencies are announced, the DM believes the announcer tends to change her action. Hence, the announcer may prefer to shroud some contingencies in order to let the DM choose the correct action, otherwise "talking-too-much" is misleading to the DM.

In next section, we will examine an application of the model.

3.1.3 Forward Induction and Insurance Contracts

The insurance problem is considered in Filiz (2006) which extends Ozbay (2006). There are two contracting parties: an insurer (he) and an insuree (she), indexed by P and A as well. There is a good of value v for the insuree. The same as before, the objective probability space is $(\Omega, \Pr)$ with $\Pr(\omega) > 0$ for all $\omega \in \Omega$. A finite subset of positive real numbers Ω represents the potential damages.[17] The insurer knows $(\Omega, \Pr)$. But the insuree is only aware of $\Omega_0 \subset \Omega$ with her subjective probability Q_0 defined on Ω_0 such that $Q_0(\omega) = \Pr(\omega|\Omega_0)$ for all $\omega \in \Omega_0$.

The insurer knows that the insuree is unaware. In contrast to last section, the insurer is not only choosing the announcement of states $V \in 2^{\Omega}$ but also proposing the amount of transfer contingent on the state and the premium. Put it differently, what the insurer proposes is a contract. Formally, we have the following definition:

Definition 3.4 *A* ***contract*** *is a triplet $C \equiv (V, t, p)$ where $V \subset \Omega$ is the insurer's announcement of states, $t : V \longmapsto \mathbb{R}_+$ is the transfer function and $p \in \mathbb{R}_+$ is the premium.*

We denote the set of all C by $\mathbb{C}$. If a state $\omega \in \Omega \setminus V$ occurs, then there is no transfer from the insurer to the insuree at all. But p has to be paid in all states. Thus, though a contract might be informationally incomplete, it is always obligatorily complete, since the obligation of each party is well-defined.

After a contract $C = (V, t, p)$ is proposed, the insuree is aware of states in $\Omega_0 \cup V$ with subjective probability $Q_C \in \Delta(\Omega_0 \cup V)$. The same as in last section, we assume that, for all $C \in \mathbb{C}$ and $\omega \in \Omega_0 \cup V$, $Q_C(\omega) > 0$ and, for all $C \in \mathbb{C}$ and $\omega \in \Omega_0$, $Q_C(\omega|\Omega_0) = \Pr(\omega|\Omega_0)$. Thus, if $V \setminus \Omega_0 = \emptyset$, $Q_C = Q_0$. That is, if no awareness updating is involved, the insuree still holds her initial belief. The strategy of insuree is a decision function $d : \mathbb{C} \longmapsto \{buy, reject\}$. That is, given the contract the insuree just decides whether to buy it.

[17] This is a reduced form of the problem. Of course, the problem can be modeled by a mapping from set of contingencies to set of damages, which would be sloppy.

Both insurer and insuree are expected utility maximizers. The insurer is risk-neutral. The expected utility of the insurer is

$$U_P(C, d(C)) \equiv \begin{cases} p - \sum_{\omega \in V} t(\omega)\,\Pr(\omega) & \text{if } d(C) = buy \\ 0 & \text{if } d(C) = reject. \end{cases}$$

The insuree is risk averse with a concave von Neuman-Morgenstein utility function u. The expected utility of the insuree is

$$U_A(C, d(C)|Q_C) \equiv \begin{cases} \sum_{\omega \in V} u(v - \omega + t(\omega) - p)Q_C(\omega) + \sum_{\omega \in \Omega_0 \setminus V} u(v - \omega - p)Q_C(\omega) \\ \text{if } d(C) = buy \\ \sum_{\omega \in \Omega_0 \cup V} u(v - \omega)Q_C(\omega) \\ \text{if } d(C) = reject. \end{cases}$$

Note that the expected utility of the insurer is an objective one while the expected utility of the insuree is subjective because of DM's unawareness. Moreover, from the insuree's perspective, the insurer's expected utility of a contract C is

$$U_P^A(C, d(C)|Q_C) \equiv \begin{cases} p - \sum_{\omega \in V} t(\omega)Q_C(\omega) & \text{if } d(C) = buy \\ 0 & \text{if } d(C) = reject. \end{cases}$$

Definition 3.5 *An equilibrium is a triplet* (C^*, d^*, Q) *such that*

(i) $C^* \in \arg\max_{C \in \mathbb{C}} U_P(C, d^*(C))$.

(ii) For all $C \subseteq \mathbb{C}$, $d^*(C) = buy$ *if* $U_A(C, buy|Q_C) \geq U_A(C, reject|Q_C)$ *and* $d^*(C) = reject$ *otherwise.*

In equilibrium, both parties are rational, since each gives the best response given the strategy of the other. It is analogous to rational assessment in last section. Mutual rationality is natural in many contractual situations.

Furthermore, we assume that $t(\omega) \leq \omega$ for all ω. Otherwise the insurer can earn an infinite profit. To show this, suppose the insuree accepts contract (V, t, p), then she must accepts contract $(V, t', p + k)$ with $t'(\omega) = t(\omega) + k$ for all ω. If $V \neq \Omega$, then the insurer can always raise his expected profit by raising k.

However, in equilibrium, the insuree's subjective probability Q is too arbitrary here. Now we try to refine the equilibrium.

Definition 3.6 *A subjective probability* Q_C *is* ***compatible*** *with respect to* $C = (V, t, p)$ *iff* $U_P^A(C, buy|Q_C) \geq 0$.

We denote the set of all compatible Q_C by Π_C. The insuree with a compatible Q_C reasons that if a contract is offered, then the insurer should not receive a negative expected profit.

Definition 3.7 *An equilibrium* (C^*, d^*, Q) *is compatible if, for all* $C \subseteq \mathbb{C}$, $d^*(C) = buy$ *implies* $\Pi_C \neq \varnothing$.

The requirement of a compatible Q_C reflects that the insuree is sophisticated in the sense that she can rule out too good to be true offers. Based on this idea, we can refine the equilibrium further.

Definition 3.8 *An equilibrium* $(C^* = (V^*, t^*, p^*), d^*, Q)$ *is consistent if, for all* $C = (V, t, p) \subseteq \mathbb{C}$ *such that* $\Omega_0 \cup V \subseteq \Omega_0 \cup V^*$, *we have* $U_P^A(C^*, d^*(C^*)|Q_{C^*}) \geq U_P^A(C, d^*(C)|Q_{C^*})$.

A consistent equilibrium requires that, from the insuree's view, the insurer provides the best offer. It is analogous to the awareness equilibrium in last section.

Filiz (2006) shows that the equilibrium with incomplete contracts exists. Moreover, she studies the case in which the insuree is ambiguity averse, and shows that competition promotes awareness of the insuree.

3.2 Partially Unforeseen Contingencies

3.2.1 Contracting with Awareness of Unawareness

> "Probability judgments are attached *not* to events but to descriptions of events."
>
> —— Amos Tversky and Derek J. Koehler (1994, p. 548)

In many contracting environments, agents cannot be aware of all the contingencies in the future. In other words, there are some unforeseen contingencies to them. Consider an insuree buys some home insurance. If the insuree is only aware of one contingency of calamity, which is "fire", she is then unaware of many other possible calamities, say explosion, earthquake, lightning, storm, flood etc.

However, agents are aware that they *may* be unaware of something. Suppose the insurer has two types of contracts for the insuree: one compensates the insuree for her loss only at the contingency of fire, and the other compensates the insuree by the same amount of money in all calamity-contingencies. If the premiums of two contracts are identical, as one would expect, the insuree prefers the later contract.[18] Even though the insuree is only aware of "fire", she is aware that there may be potentially many unforeseen calamity-contingencies. The insuree can therefore classify the contingencies based on a general concept: "calamity". Although the insuree cannot spell out the residual unforeseen contingencies in the event "calamity", she is aware of the possibility that those contingencies exist.

Moreover, if the insurer describes the latter contract differently, say replacing the general term "calamity" by "fire, explosion, earthquake, lightning, storm, flood or other calamities", the insuree will be aware of those extra contingencies. If the insuree underestimates the existence of those contingencies when only "calamity" is mentioned, she is willing to pay more for the insurance after being more aware. Thus when the contingencies in the contract are framed differently, the insuree's preference over the same alternative (the contract or her outside option) is manipulated. This is a typical result of framing effect in contracting, which the standard contract theory abstracts from.

In the paper, we consider a bilateral contracting problem to explore how the principal (he) frames the contingencies in the optimal contract against the agent (she) who is unaware but aware of her unawareness. In contrast to the standard contract theory where the contract is reduced to a mapping from the realized contingency to the actions, the principal here additionally decides how to frame the contingencies in the contract, and contemplates whether or not to make the agent more aware. In the general setting, we use σ-algebra to model the richness of the agent's language. The agent can assign probability to an event only in her language. We call the contract vague if the agent is still unaware of some payoff-relevant contingencies after the principal proposes the contract. We show that the optimal contract is vague if and only if the principal exploits the agent.

The focus on the paper are the general framework for analyzing problems where the more aware principal contracts with the less aware agents, and the explanation for vague terms in contracts. Moreover, we use the model to discuss several particular problems.

In an insurance problem, the insurer makes the unaware insuree fully aware if the insuree is aware that she is unaware of some unforeseen calamities and slightly underestimates their existence. Conversely, the insurer is silent on the insuree's unforeseen calamities. In one case, suppose the insuree underestimates the unforeseen calamities too much. The insurer then obtains a higher profit by providing low benefits in her unforeseen contingencies. In the other case, suppose the insuree overestimates the unforeseen calamities. The insurer benefits from raising both premium and benefit for the insuree in her un-

[18] Suppose the legal term "calamity" that covers "fire" is well-defined and verifiable. Furthermore, the insurance contract is perfectly enforced.

foreseen contingencies. However, the unforeseen contingencies are not so likely to occur. In both cases, the insurer exploits the insuree. Thus only a certain range of degrees of awareness of unawareness prevents the insuree from exploitation. Sometimes awareness of unawareness is valuable.

In a contracting problem with force majeure clauses, i.e., clauses which free some party from obligation when an unforeseen circumstance beyond the control of the parties occurs, such as war, strike, riot, crime, act of God (e.g., fire, flood etc.), we show that it is always optimal for the employer to propose a vague contract with a general term "force majeure" but not to describe the particular unforeseen contingencies, which promote the contractor's awareness, no matter how aware of her unawareness the contractor is. If the contractor underestimates the existence of the force majeure, the employer charges her for a higher transfer in the force majeure event. Conversely, the employer charges her for a higher transfer in the non force majeure event. In both cases, the extra transfer is more likely to occur than the contractor believes. Since the contractor is always exploited by the employer, the policy recommendation is promoting the contractor's awareness of the particular force majeure before contracting. The following suggestions are from Liblicense on the web:

> "To make sure that the parties know exactly what is and is not a legitimate excuse for failure to provide access to licensed materials, it would be better to specifically set forth the circumstances that excuse a failure of performance, rather than rely on a general force majeure clause."[19]

Then, we illustrate the persuasive advertising result of an experience good. We show that the firm has an incentive to make the consumer only aware of the good contingency of consumption if and only if the consumer underestimates both good and bad experiences. Because the advertisement raises the consumer's subjective valuation of the good, this is exactly the persuasive advertising result. However, the insuree's belief is wrong. Since the consumer puts too much weight on the good contingency, she is exploited in the objective world. In this sense, persuasive advertising of experience goods is exploitative.

Lastly, we show that a benevolent parent frames the contingencies in the future for his kid, and thus manipulates the kid's belief. This makes the kid more optimistic. The kid therefore overcomes her self-control problem.

Related Literature:

Psychology:

Tversky and Kahneman (1974) originate the research on human judgment of probability for descriptive purpose in science. They argue that people use several heuristics to assess

[19] Liblicense: Licensing Digital Information (http://www.library.yale.edu∼llicense/forcecls.shtml).

probability, and one of them is *availability*.[20] Tversky and Kahneman (1974, p.1127) argue:

> "There are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind".

In peoples' minds, the probability judgment of an event depends on whether its instances can be retrieved. The availability of some instances is equivalent to awareness of some contingencies in our model. If a contingency is not available to the agent, we say the agent is unaware of the contingency, or the contingency is unforeseen by the agent.

A more relevant work is *support theory* by Tversky and Koehler (1994). They introduce an alternative theory of subjective probability that deviates from the Bayesian model by withdrawing the additivity of probability measure. The judged probability is modeled by the relative support values of the focal and alternative hypotheses. Empirical evidences suggest that the support function is *subadditive* for implicit disjunctions, that is, the probability judgment of an implicitly disjunctive event is smaller than the probability judgment of the same but explicitly unpacked event. One of the reasons for it is that unpacking an event enhances the availability of particular contingencies in the event. It shares a similar idea with the present model when the agent underestimates the existence of potential unforeseen contingencies, although there is some difference between these two approaches as we see below.[21]

Concerning the insurance-purchasing decision, Johnson et al. (1993) present some questionnaire evidences to show that illustration of vivid calamities increases the insuree's valuation of the insurance. In one application of the present paper, we explore systematically the insurance problem based on these psychological effects. However, we show that announcing vivid calamities is not always optimal for the insurer.

Modeling Unforeseen Contingencies:

Roughly speaking, the agent fails to foresee some event if she has not thought about it when she makes a decision. In economics, there are two main approaches to model unforeseen contingencies: decision-theoretic approach and epistemic approach. (See a survey by Dekel et al. (1998a))

[20] See also the original paper on availability by Kahneman and Tversky (1973).

[21] Although, in a subsequent work, Rottenstreich and Tversky (1997) show that subadditivity is also valid for explicit disjunctions, the present paper abstracts from this effect and focuses only on implicit subadditivity. Our motivation is that unpacking of an implicitly described events can update peoples' awareness of some relevant contingencies, and is therefore more relevant to unforeseen contingencies in contracting problems.

Decision theoretic approach starts from the agent's preference or choice behavior without referring to the agent's true belief. The most relevant paper is Ahn and Ergin (2007). Unforeseen contingencies are modeled by generalizing standard subjective expected utility theory through partition-dependent framing effects. Ahn and Ergin (2007) also provide an axiomatic foundation of a generalized version of support theory introduced above. In spite of the relevance to our work, our paper cannot be captured by their model. For instance, in the insurance example in section 3.2.6, the agent's subjective probability for a vague contract and that for a non-vague contract are different, since there is a difference between announcement of the particular contingency "flood" and saying "other calamities". But, in Ahn and Ergin (2007), these two contracts have no difference for the agent, since the partitions of the set of contingencies in these two contracts are identical. Thus modeling unforeseen contingencies by partition-dependent framing effects loses some important considerations. Therefore, developing the decision theoretic foundation for the present model should be important for the future research.

In contrast, epistemic approach starts from the agent's belief. It directly models the knowledge of an event per se as a distinct event. If the agent fails to foresee an event, we say she is *unaware* of the event. Modica and Rustichini (1994) first study unawareness by epistemic approach. Modica, Rustichini and Tallon (1998) applies unawareness to a general equilibrium model to explain bankruptcy. Later, Heifetz et al. (2006) and Li (2009) independently model unawareness that circumvent the impossibility result of non-trivial unawareness by Dekel et al. (1998b). Thus U (Unawareness) is possible to express. Considerable progress has been made such that Awareness of Unawareness (AU) is possible to be expressed. (See, e.g., Board and Chung, 2007) AU plays a role in our model. Agents are unaware of some future contingencies, while they are aware that they may be unaware of something. This changes the contracting result in many important aspects. For example, in our paper, an insuree is aware that there may be many potential unforeseen calamities. An appropriate degree of AU refrains the insuree from exploitation by the insurer.

Games with Unawareness:

Recently, many papers study games with unawareness. We only discuss those papers that are very relevant to our work. Ozbay (2008), and fundamentally Heifetz et al. (2008), studies strategic announcement of some contingencies. The difference from our work is on the agent's subjective probability of the newly announced contingencies. We assume that the agent can put correct weights on all contingencies she is aware of due to her ability to judge the frequencies of all vivid events. Furthermore, we do not require that the agent accepts only a justifiable contract that requires that agent's cognitive ability to reason the principal's profitability, as implicitly assumed in most bounded rationality literature. But our paper can be captured by Halpern and Rego (2006). Halpern and

Rego (2006) provide a general setting for studying games with unawareness of actions (possibly the actions of the nature). AU of the agent is modeled by allowing some player to make a "virtual move". In our paper, although the agent cannot be aware of all particular contingencies in the general event, she believes that the nature can make some virtual move based on her subjective probability.

Unawareness and Contract Design:

Firstly, there are some papers on unawareness of endogenous variables, say actions of some contracting parties. Gabaix and Laibson (2006) study how the firm exploits the consumers who are unaware of later add-on prices. Zhao (2008) introduces unawareness into moral hazard problem, and analyses the value of awareness of additional actions. von Thadden and Zhao (2009) provide incentive design for an agent who is unaware of some choice possibilities.

Secondly, there are also some papers on unawareness of exogenous variables, say actions of nature (contingencies). Our paper belongs to this category. Besides it, Filiz-Ozbay (2008) incorporates unawareness into insurance contracts. Chung and Fortnow (2007) model a two-stage game of interaction between a contract (or law) writer and an interpreter. In Tirole (2009), a buyer is aware that the design sold by a seller may not be appropriate, and therefore invests some cognitive resources on thinking whether or not she is indeed unaware of something.

Other non-Bayesian Reasoning Models:

The paper belongs to the growing literature on interaction between a fully rational principal and a boundedly rational agent who uses a non-Bayesian learning rule. von Thadden (1992) studies a repeated contracting problem between a seller and a buyer who uses a non-strategic learning rule. Given the rule, in the long run, the buyer is free from exploitation. Piccione and Rubinstein (2003) model differences among consumers in their ability to perceive intertemporal patterns of prices. Spiegler (2006) shows that the patients using anecdotal reasoning suffer from the exploitation by quacks. Shapiro (2006) studies how a firm manipulates a consumer's memory of the consumption experience when consumers have imperfect recall. Mullainathan et al. (2007) discuss the principal's persuasion method by metaphor when the agent puts uncorrelated situations into one category.

The plan of the rest of the paper is as follows: In section 4.3.2, we provide the general model of framing contingencies in contracts in full details. Section 3.2.6 applies the model in an insurance problem. Section 3.2.7 presents a contracting problem with force majeure clauses. Section 4.3 discusses the persuasive advertising. Section 3.2.9 uses the model to view self-control problems. The last section concludes. For the ease of exposition, we put all the proofs in the appendix.

3.2.2 Model

3.2.3 Language and Contracts

There are two parties involved in the contracting situation: a *principal* (he) P and an *agent* (she) A. The principal proposes a contract to the agent. The agent decides whether to accept it.

We assume that the principal is omniscient. He knows everything that the analyst knows. This assumption is strong but still plausible in situations where the principal is an experienced firm with many experts, whereas the agent is a naive individual (a consumer or an employee) who lacks sufficient contracting experience.

Let Ω denote a finite set of contingencies consisting of exclusive and exhaustive elements ω.

We assume that the agent is aware of only some contingencies in Ω. Let $K_0(\subset \Omega)$ denote the set of contingencies, which the agent is aware of. We call K_0 the agent's awareness. In terms of psychology, the elements in K_0 are the only available concrete scenarios in the agent's mind.

$X(\subset \Omega)$ represents a non-empty *general* event that is determined by a generic characteristic of contingencies. The characteristic leads to a dichotomic classification of payoff-relevant contingencies for the agent. Put it differently, X captures a general concept the agent understands, no matter whether or not the individual elements in X are available in the agent's mind. Similarly, its complement $X^C \neq \emptyset$ is also a general event. Both X and X^C are payoff-relevant to the agent. The economic meaning of this general event is captured by the agent's utility function, as we shall see later. Although it is more realistic to assume many general events based on other characteristics of contingencies, here we only focus on the most payoff-relevant one in the context under consideration, say "calamity" event for the insuree, or "good-experience" event for the traveler. Example 3.1 illustrates X and K_0 intuitively.

Example 3.1 *Let the set of contingencies be $\Omega = \{no\ calamity, fire, flood, earthquake\}$. $X = \{fire, flood, earthquake\}$ is a general event: "calamity". All calamity contingencies share the same characteristic that the insuree loses her assets, safety or health in these contingencies. However, the insuree is not necessarily able to list all the contingencies in X. Let $K_0 = \{no\ calamity, fire\}$, thus the agent is unaware of flood and earthquake. Figure 3.1 depicts X and K_0 graphically.*

As "probability judgments are attached *not* to events but to descriptions of events" (Tversky and Koehler, 1994, p.548), we define the language for the agent to express

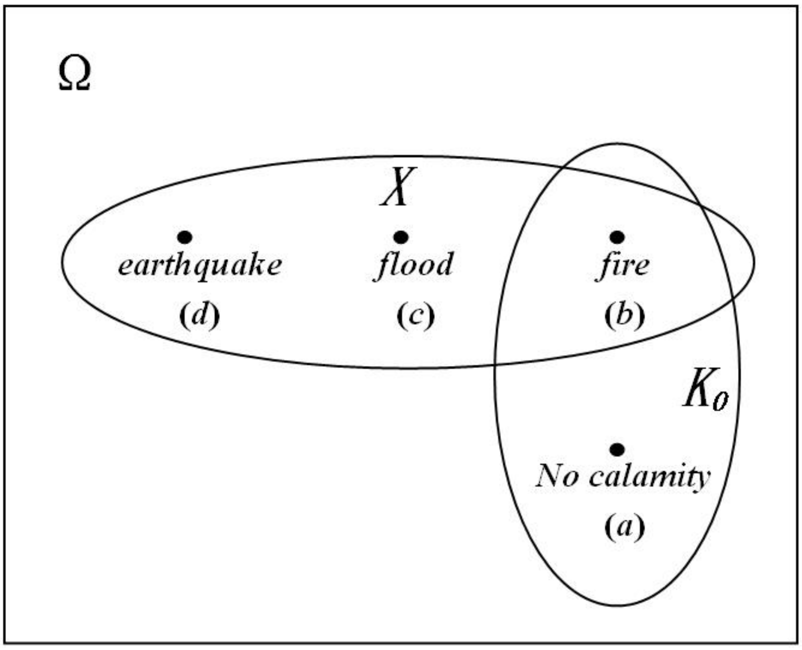


Figure 3.1: X and K_0 in Example 3.1

events in a contract, given the set of contingencies Ω, a general event X and the agent's awareness K_0.

Definition 3.9 *The* ***language*** *of the agent with awareness K_0 is $\mathcal{L}(K_0)$ that is the smallest ($\sigma-$) algebra over Ω such that*[22]

1. *$X \in \mathcal{L}(K_0)$ and*

2. *For all $\omega \in K_0$, we have $\{\omega\} \in \mathcal{L}(K_0)$.*

If an event is in $\mathcal{L}(K_0)$, we say the event is *expressible* for an agent with awareness K_0. Property 1 reflects, although the agent may be unaware of some contingencies in X, she can express the general event X simply by an abstract term, say "calamity". Property 2 says, since the agent is aware of each contingency in K_0, she can express each singleton event $\{\omega\} \subseteq K_0$. Since Ω is finite, there is no difference between σ-algebra and algebra here. $\mathcal{L}(K_0)$ is closed under complements, intersections, and unions, which represent "not", "and", "or" in natural language. The set of expressible events is K_0-dependent. The larger the set K_0, the richer the σ-algebra. In words, the awareness of the agent determines the richness of her language. An example of $\mathcal{L}(K_0)$ is shown in Example 3.2.

[22]There is no difference between $\sigma-$algebra and algebra here, since Ω is finite.

Example 3.2 *Based on Example 3.1, for brevity, let* $a \equiv$ *no calamity,* $b \equiv$ *fire,* $c \equiv$ *flood,* $d \equiv$ *earthquake. We have* $\Omega = \{a, b, c, d\}$*,* $X = \{b, c, d\}$*, and* $K_0 = \{a, b\}$*.*

The agent's language is thus $\mathcal{L}(K_0) = \{\emptyset, \{a\}, \{b\}, \{c, d\}, \{a, b\}, \{a, c, d\}, \{b, c, d\}, \Omega\}$*, that is the collection of all events the agent can express. For instance, the agent can express the event* $\{a, c, d\}$ *as "no calamities or calamities but not fire". However, say, the event* $\{c\}$ *is not expressible. To express it, the agent has to be aware of* c *or* d*.*

In standard contract theory, a contract is reduced to a mapping $C : \Pi \mapsto S$ where Π is a partition of Ω and S is the choice set of the two parties.[23] The partitional domain of C is tantamount to the case where C is a complete contract. Suppose that Π is not a partition. If $\cup_{E\in\Pi} E \neq \Omega$, there are gaps in C. (See Shavell, 2006) If $E \cap F \neq \emptyset$ for some E, $F \in \Pi$, there may be contradictions in C. (See Heller and Spiegler, 2008) In this paper, we focus only on complete contracts. Given the agent's language $\mathcal{L}(K_0)$, we can now model how the agent uses her language to form a contract.

Since, in general, not all events are expressible by the agent with awareness K_0, the partition Π is not arbitrary. Let $\Pi(K_0)$ denote the *finest* partition of Ω with respect to $\mathcal{L}(K_0)$. Formally, $\Pi(K_0)$ is a partition of Ω such that $E \in \mathcal{L}(K_0)$ for all $E \in \Pi(K_0)$ and there is no $E' \subset E \in \Pi(K_0)$ and $E' \neq \emptyset$ such that $E' \in \mathcal{L}(K_0)$. The following lemma explicitly describes the finest partition $\Pi(K_0)$.

Lemma 3.2 $\Pi(K_0) = \{\{\omega\} : \omega \in K_0\} \cup \{X \setminus K_0\} \cup \{X^C \setminus K_0\}$.

Proof 3.4 *See Appendix A.1.2.*

Lemma 3.2 shows the finest partition that the agent with awareness K_0 can express is the collection of all singleton events the agent is aware of and two residual unforeseen general events $X \setminus K_0$ and $X^C \setminus K_0$.

Using the finest partition $\Pi(K_0)$, we define the contract within the agent's awareness K_0 as follows:

Definition 3.10 *A* ***contract*** *with* K_0 *is a mapping* $C^{K_0} : \Pi(K_0) \mapsto S$*.*

The contract maps from the finest expressible event to their choice.

Example 3.3 *Based on Example 3.2, we have that* $\Pi(K_0) = \{\{a\}, \{b\}, \{c, d\}\}$ *is the finest partition. The contract with* K_0 *is*

$$C^{K_0} = \{(\{a\}, s_1), (\{b\}, s_2), (\{c, d\}, s_3)\} = \begin{pmatrix} \textit{no calamity} & \rightarrow s_1 \\ \textit{fire} & \rightarrow s_2 \\ \textit{other calamities} & \rightarrow s_3 \end{pmatrix},$$

[23] Π is a partition of Ω if $\cup_{E\in\Pi} E = \Omega$ and $E \cap F = \emptyset$ for all E, $F \in \Pi$.

where s_i represents a particular choice.

Of course, the principal can also announce some contingency out of K_0 to enlarge the agent's awareness. We will return to this point in section 3.2.5.

3.2.4 Probabilities

We define the *objective* probability space by $(\Omega, 2^\Omega, \mu)$ where μ is the objective probability measure on 2^Ω, which is the collection of all subsets of Ω. Since the principal is omniscient, he knows $(\Omega, 2^\Omega, \mu)$. Assume that $\mu(\{\omega\}) > 0$ for all $\omega \in \Omega$, so no contingency is trivially impossible.

However, the agent is unaware. Due to her limited language $\mathcal{L}(K_0) \subseteq 2^\Omega$, she is unable to judge the probability of all events in Ω. Moreover, her probability judgment of her expressible event may also be wrong, as a result of her unawareness. However, it is innocuous to assume that the agent has correct relative weights of the contingencies in K_0. Intuitively, the agent can objectively judge the frequency of each contingency within her awareness. For example, the agent has access to some data services, which gives her the frequencies as long as she puts in an explicit inquiry of the contingencies in her mind. Equivalently, by reducing one degree of freedom, we assume that the agent knows $\mu(\{\omega\})$ for all $\omega \in K_0$. In Example 3.1, the insuree is able to judge the frequency of “fire”. Since the insuree is aware of fire, she can use some device, say Internet, to acquire the information.

However, the agent has her *subjective* weights on two residual unforeseen general events $X \setminus K_0$ and $X^C \setminus K_0$. For $Z \in \{X, X^C\}$, let $\alpha_Z(K_0)$, which is known by the principal, be a non-negative weight of the agent's residual unforeseen event in Z.[24] In other words, $\alpha_Z(K_0)$ represents the agent's degree of *awareness of unawareness* (AU) of unforeseen contingencies in Z. In Example 3.1, the insuree is aware that there may be some other calamities with her subjective frequency $\alpha_X(K_0)$, although the insuree cannot tell what they are exactly.

We make the following assumption on $\alpha_Z(\cdot)$.

Assumption 3.2 *If $Z \subseteq K_0$, then $\alpha_Z(K_0) = 0$ for $Z \in \{X, X^C\}$.*

Assumption 3.2 reflects that if the agent is aware of every contingency $\omega \in Z$, then she has a correct belief that there are no residual unforeseen contingencies in Z. It

[24] The conjunction fallacy by Tversky and Kahneman (1983) shows that α_Z can be negative if the smaller event $Z \cap K_0$ is not available when the agent judges the probability of the larger event Z. However, since we have assumed that the agent is fully aware of contingencies in K_0, all contingencies in $Z \cap K_0$ are available to the agent. Thus $\alpha_Z(K_0) < 0$ is ruled out.

is natural that full awareness of contingencies in a general event implies no unforeseen contingencies in the event.

Thus from the agent's view, the probability space is $(\Omega, \mathcal{L}(K_0), \mu^{K_0})$. The agent's *subjective* probability measure is $\mu^{K_0} : \mathcal{L}(K_0) \mapsto \mathbb{R}_+$ such that

$$\mu^{K_0}(\{\omega\}) \equiv \frac{\mu(\{\omega\})}{\sum\limits_{\omega \in K_0} \mu(\{\omega\}) + \sum\limits_{Z \in \{X, X^C\}} \alpha_Z(K_0)} \text{ for all } \omega \in K_0 \text{ and}$$

$$\mu^{K_0}(Z \setminus K_0) \equiv \frac{\alpha_Z(K_0)}{\sum\limits_{\omega \in K_0} \mu(\{\omega\}) + \sum\limits_{Z \in \{X, X^C\}} \alpha_Z(K_0)} \text{ for all } Z \in \{X, X^C\}.$$

The agent can only assign probabilities to events within her language, and is able to assign probabilities to all events in her language. Furthermore, the agent's subjective probability measure depends on her awareness K_0. The agent uses a heuristic to judge the probability. She assigns a weight $\mu(\{\omega\})$ to each contingency in K_0. If the agent is aware of ω, that is, $\omega \in K_0$, then her probability judgment of $\{\omega\}$ is the ratio of the weight of ω to the sum of the weights of all foreseen contingencies and the residual unforeseen events. The agent's probability judgment of the residual unforeseen event is the ratio of the weight of the unforeseen event to the sum of all weights. The way how the agent forms probability is similar to support theory initiated by Tversky and Koehler (1994) and developed by Ahn and Ergin (2007). The difference has been discussed in the introduction.

Obviously, by Assumption 3.2, if $K_0 = \Omega$, then the agent's subjective probability measure is nothing but the objective one, that is, $\mu^{K_0} = \mu$.

If $\alpha_Z(K_0) = 0$, then $Z \setminus K_0$ is a completely unforeseen event. Since the agent gives zero weight to the residual event $Z \setminus K_0$, the agent believes that she is aware of every contingency in Z. Most applied unawareness papers are in this case where the agent is unaware and unaware of her unawareness. In Example 3.1, the insuree believes that fire is the mere calamity in this case.

If $\alpha_Z(K_0) = \sum\limits_{\omega \in Z \setminus K_0} \mu(\{\omega\})$, then the agent has a correct belief of the weight to the residual unforeseen event $Z \setminus K_0$. The agent's degree of AU makes her behave *as if* she foresees $Z \setminus K_0$, although she cannot explicitly express the particular contingencies in $Z \setminus K_0$. This case is degenerate to the situation in which the agent only cannot describe the contingencies in $Z \setminus K_0$ as in Maskin and Tirole (1999) and Tirole (1999). In Example 3.1, the insuree is unaware of flood and earthquake, but she has a correct belief of the probability of the event that other calamities occur.

If $\alpha_Z(K_0) < \sum\limits_{\omega \in Z \setminus K_0} \mu(\{\omega\})$, then the agent is aware that there may be potentially other contingencies in $Z \setminus K_0$, but *underestimates* their existence.[25]

Finally, if $\alpha_Z(K_0) > \sum_{\omega \in Z \setminus K_0} \mu(\{\omega\})$, the agent *overestimates* the existence of potential other contingencies in $Z \setminus K_0$.

The interpretation of different values of $\alpha_Z(K_0)$ is depicted in Figure 3.2.

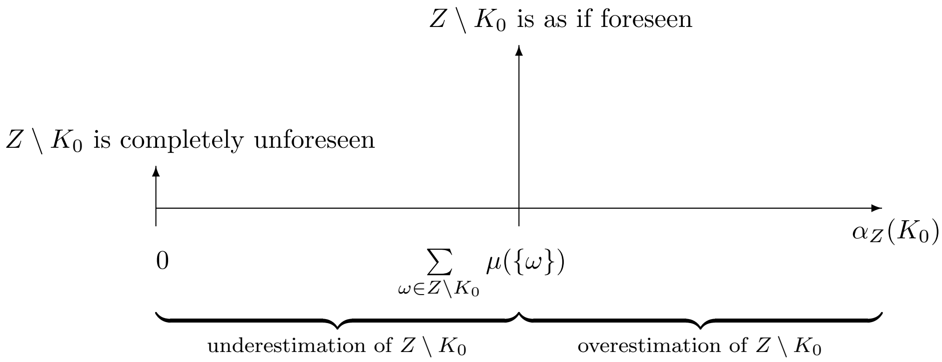


Figure 3.2: Interpretation of $\alpha_Z(K_0)$ with $Z \in \{X, X^C\}$

3.2.5 Framing Contracts

In standard contract theory, we can reduce any contract to a mapping $C : \Omega \mapsto S$. However, it abstracts from the details of how the event in each clause is described. We consider the following two contracts following Example 3.1:

$$C_1 = \begin{pmatrix} \text{no calamity}, & s_1 \\ \text{fire}, & s_2 \\ \text{other calamities}, & s_3 \end{pmatrix}, C_2 = \begin{pmatrix} \text{no calamity}, & s_1 \\ \text{fire}, & s_2 \\ \text{flood}, & s_3 \\ \text{earthquake}, & s_3 \end{pmatrix}.$$

Although two contracts represent the same reduced mapping, they are framed[26] differently. C_2 makes the agent additionally aware of *flood* and *earthquake*. Of course, in reality, besides *flood* and *earthquake* there are many other calamities. Then saying

[25] $\alpha_Z(K_0) \leq \sum_{\omega \in Z \setminus K_0} \mu(\{\omega\})$ with $Z \in \{X, X^C\}$ is nothing but the result of subadditivity of implicit disjunction by Tversky and Koehler (1994).

[26] In general, framing effect says people's perception of an object will be different if the object is put into a different context, or described differently. Here, two insurance contracts have the same underlying mapping, but the insuree's preference is distorted by a different description of contingencies.

"other calamities" and "'flood or earthquake" are indeed different. In essence, we distinguish only two options here: expressing a general term of the event and listing all individual contingencies in this event.

In general, the principal can update the agent's awareness K_0 via the contract. The agent becomes aware of $\omega \notin K_0$ if ω is explicitly announced in the contract. Formally, we denote the awareness of the agent after reading the contract by K ($\supseteq K_0$). K consists of the contingencies which the agent is aware of after understanding the contract. Thus the principal chooses the *framing* K such that the agent's language becomes $\mathcal{L}(K)$, and the finest partition is refined to $\Pi(K)$. The principal can therefore enrich the language of the agent by framing contingencies.

Furthermore, the agent's subjective weights of the residual unforeseen events become $\alpha_Z(K)$. Roughly speaking, the new contingencies in the agent's mind change the agent's conjectural amount of the residual unforeseen contingencies. Consequentially, the agent's subjective probability measure becomes μ^K. This is not a standard Bayesian updating, since what is updated here is the probability space $(\Omega, \mathcal{L}(K), \mu^K)$ as a whole.

In contrast to the approach of biased belief about contingencies, the biased belief of the agent here is derived purely from the agent's awareness. More importantly, the principal can adjust the agent's biased belief only according to the way how the agent's awareness is updated.

Definition 3.11 *We call a contract C^K* ***vague*** *in Z if $Z \nsubseteq K$ where $Z \in \{X, X^C\}$.*

In other words, a contract C^K is vague in Z if the agent is still unaware of some contingencies in Z after C^K is proposed. We say a contract C^K is *vague* if C^K is either vague in X *or* in X^C. Moreover, we say a contract C^K is *less vague* than $C^{K'}$ if $K \supseteq K'$.[27]

Let the agent's von Neumann-Morgenstern (v.N.M) utility function be $u_A : \Omega \times S \mapsto \mathbb{R}$ such that

$$u_A(\omega, s) \equiv \begin{cases} u_A^X(s) & \text{for } \omega \in X \\ u_A^{X^C}(s) & \text{for } \omega \in X^C \end{cases} .$$

For $Z \in \{X, X^C\}$, $u_A^Z(s)$ represents the agent's utility level of choice s when a contingency $\omega \in Z$ occurs. The agent's v.N.M utility function is contingency-dependent. But it depends only on whether the contingency falls into X or not. Put differently, the utility function is "general-event-dependent". The difference between u_A^X and $u_A^{X^C}$ captures exactly the *economic meaning* of the general event X. For example, let X denote

[27] In contrast to contractual incompleteness that is a concept independent of the agents' awareness, the concept of vagueness depends on the agent's initial awareness K_0. For example, if $K_0 = \Omega$ (the agent is fully aware before contracting), then the contract can never be vague, because X, $X^C \subseteq K_0$.

the calamity event. $u_A^X(s) \neq u_A^{X^C}(s)$ reflects that the agent's utility of s in calamity is different from the utility of the same choice s when no calamity occurs. We also assume that the utilities of s are the same within X or X^C. In other words, if calamity happens the agent feels bad to the same extent, no matter it is a fire or a flood.

The agent's subjective expected utility of a contract C^K is therefore

$$\sum_{E\in\Pi(K)} \mu^K(E) \sum_{Z\in\{X,X^C\}} I_{E\subseteq Z} \cdot u_A^Z(C^K(E))$$

where $I_{E\subseteq Z}$ is an index function. If $E \subseteq Z$, we have $I_{E\subseteq Z} = 1$. Otherwise, $I_{E\subseteq Z} = 0$.

We denote the principal's v.N.M utility function by $u_P : S \mapsto \mathbb{R}$ that is contingency-independent. After all, whether or not the contingency s falls into the general event X is only payoff-relevant for the agent. Thus the principal's (objective) expected utility of C^K is

$$\sum_{E\in\Pi(K)} \mu(E)u_P(C^K(E)).$$

Since different problems have different restrictions on the choice of contracts, we denote the set of all *admissible* contracts with framing K by $\mathbb{C}^K$. The particular specification of $\mathbb{C}^K$ depends on the particular context under consideration.

The problem for the principal is therefore to design the optimal contract C^K, which includes the optimal framing K, subject to the agent's participation. It can be written formally as

$$\max_{K\supseteq K_0,\ C^K\in\mathbb{C}^K} \sum_{E\in\Pi(K)} \mu(E)u_P(C^K(E)) \tag{3.11}$$
$$\text{s.t.} \sum_{E\in\Pi(K)} \mu^K(E) \sum_{Z\in\{X,X^C\}} I_{E\subseteq Z} \cdot u_A^Z(C^K(E)) \geq \sum_{E\in\Pi(K)} \mu^K(E) \sum_{Z\in\{X,X^C\}} I_{E\subseteq Z} \cdot u_A^Z(\overline{s}).$$

On the right hand side of participation constraint of problem (3.11), $\overline{s}$ is the agent's outside option. Rejecting C^K means that the agent chooses $\overline{s}$ in each contingency. Since the expectation is determined by the agent's subjective probability $\mu^K(\cdot)$, the principal can also influence the agent's perception of her valuation of the outside option by choosing K.

We implicitly assume that the agent has no cognitive ability to infer the set of contingencies from the optimal contract. Her understanding of the set of contingencies is influenced only by the framing K. Thus we rule out the possibility that the agent can do the forward induction as in Heifetz et al. (2008).

The larger K, the richer the language the principal can use to express the events in the contract. However, the agent's subjective probability may be distorted in a direction that the principal dislikes. It brings us the general idea of a trade-off for choosing the optimal framing.

It is worth mentioning that it would be also interesting to study the problems richer than this two-stage game. However, different fields have different interesting considerations. At this given stage of the literature in the general contracting problem, we are restricted in this two-stage game benchmark. Richer models in the particular fields are worthy for the future research.

Slightly abusing notations, let $C^K(\omega) \equiv C^K(E)$ where $\omega \in E \in \Pi(K)$.

Definition 3.12 *We call a contract C^K* ***exploitative*** *if* $\sum_{\omega\in\Omega} \mu(\{\omega\})u_A(\omega, C^K(\omega)) < \sum_{\omega\in\Omega} \mu(\{\omega\})u_A(\omega, \overline{s})$.

In words, a contract C^K is exploitative if the agent's objective expected utility of C^K is lower than the objective expected utility of her outside option. Thus the judgment whether the agent is exploited or not is in terms of the objective probability, yet not the agent's subjective one.

To make things interesting, we make two additional assumptions:

Assumption 3.3 *If $C^K \in \mathbb{C}^K$, $K' \supseteq K$ and $C^K(\omega) = C^{K'}(\omega)$ for all ω, then $C^{K'} \in \mathbb{C}^{K'}$.*

Assumption 3.3 allows some natural flexibility on the set of admissible contracts. It says that if a contract is admissible, then any contract with a refined partition that has the same reduced mapping from the set of contingencies to the action space is also admissible. Put another way, for any vague contract that is admissible, the principal can also write a non-vague contract that shares the same reduced mapping.

Assumption 3.4 *The principal's tie-breaking rule is choosing one of the least vague C^K among the optimal contracts.*

In words, Assumption 3.4 says, whenever the principal is indifferent between making the agent more aware and being silent, the principal prefers the former. In most problems, since such tie-breaking situation is not generic, we can ignore it. Nevertheless, Assumption 3.4 is plausible in reality, because a less vague contract signals the principal's honest, specialty in his field. The principal has no incentive to shroud some contingencies unless he has a rent of doing so.

Given Assumption 3.2-3.4, we have the following proposition:

Proposition 3.3 *If C^K is an optimal contract, C^K is exploitative if and only if C^K is vague.*

Proof 3.5 *See Appendix A.1.3.*

Proposition 3.3 provides a necessary and sufficient condition that the principal exploits the agent. Hence, whenever there is a vague term in the contract, the agent must be exploited. Conversely, if the agent is exploited, the contract, which she accepted, must be vague. This proposition will be frequently used in the following sections.

The intuition of the "only if" part is straightforward. If principal exploits the agent, the participation constraint must be violated. This outcome cannot occur when the contract is non-vague (by Assumption 3.2). The intuition of the "if" part is as follows. Suppose an optimal contract is not exploitative. Then the agent will accept the contract when she is fully aware. By the tie-breaking rule in Assumption 3.4, the principal will choose a non-vague contract. Note that he principal is always able to do so due to Assumption 3.3.

3.2.6 Insurance Contracts

We consider a home insurance problem where an insurer as the principal proposes a contract to an insuree as the agent. Suppose the set of contingencies is $\Omega = \{a, b, c\}$. For simplicity, we assume there are only two contingencies of calamity: a and b. a is the contingency of "fire", and b is the contingency of "flood". The general event "calamity" is $X = \{a, b\}$, which is verifiable. If the event calamity occurs, it is either a fire or a flood. The residual contingency c is the contingency of "no calamity". Let the probability measure be $\mu(\{a\}) = p$, $\mu(\{b\}) = 1 - p - q$ and $\mu(\{c\}) = q$.

Before contracting, the insuree is fully aware of contingencies a and c while she is unaware of contingency b, that is, $K_0 = \{a, c\}$. But she is aware that there may be some other potential calamity contingencies of which she is unaware.

By Assumption 3.2, $\alpha_X(K) = 0$ for $b \in K$. In words, if b is announced in the contract C^K, the insuree will be fully aware of b and then correctly believes that there are no unforeseen calamities. In this case, the insuree understands the objective probability space $(\Omega, 2^\Omega, \mu)$ and assigns a correct probability to each event.

On the other hand, if b is not announced in the contract C^K, the insuree remains unaware of b. Let $\alpha_X(K) \equiv \alpha$ for $b \notin K$. α measures the insuree's degree of AU. She assigns a weight α to the unforeseen event $\{b\}$ while assigning weights p to a and q to c, respectively. Thus her subjective probability of b is $\frac{\alpha}{\alpha+p+q}$ and her subjective probability of a and c are $\frac{p}{\alpha+p+q}$ and $\frac{q}{\alpha+p+q}$ respectively.[28]

By definition of subjective weights, we have $\alpha \geq 0$.

If $\alpha = 0$, then $\{b\}$ is completely unforeseen by the insuree. The insuree is extremely overconfident that she regards the set of contingencies as $\{a, c\}$ where a and c have probability $\frac{p}{p+q}$ and $\frac{q}{p+q}$ respectively. In contrast, $\alpha > 0$ captures the fact that the insuree is aware that she may be unaware of some other potential calamity contingencies.

If $\alpha = 1 - p - q$, then the insuree has a correct probability judgment. Everything is as if the insuree foresees $\{b\}$, although she cannot explicitly state "flood". If $0 < \alpha < 1 - p - q$, then the insuree is aware that she is unaware of something but underestimates their existence. On the other hand, if $\alpha > 1 - p - q$, the insuree overestimates the existence of potential other calamities.

The choice of the insurer in each contingency is $t \in \mathbb{R}$. t denotes a monetary transfer from the insuree to the insurer.[29] Let the monetary value of the house be $w_1 > 0$. If there is a calamity the value of the house reduces to $w_0 \in (0, w_1)$. Assume that the insurer is risk neutral and the insuree is risk averse. The v.N.M utility function of the insuree is $u(\cdot)$ over money where $u(\cdot)$ is a smooth, strictly increasing and strictly concave function, and satisfies Inada conditions ($u'(0) = \infty$ and $u'(\infty) = 0$).

Thus we have that the insuree's v.N.M utility function is

$$u_A(\omega, t) \equiv \begin{cases} u(w_1 - t) & \text{for } \omega = c \\ u(w_0 - t) & \text{otherwise} \end{cases}. \tag{3.12}$$

The insurer's v.N.M utility is $u_P(t) = t$.

There is no restriction on the insurer's choice t in each contingency. Thus the set of admissible contracts with K is the set of all contracts. Assumption 3.3 is therefore satisfied.

Case 1: Non-Vague Contracts

Firstly, we consider that the insurer proposes a contract where b is announced:

$$C^{\{a,b,c\}} = \begin{pmatrix} a, & t_a \\ b, & t_b \\ c, & t_c \end{pmatrix} = \begin{pmatrix} \text{fire}, & t_a \\ \text{flood}, & t_b \\ \text{no calamity}, & t_c \end{pmatrix}.$$

[28] More precisely, the insuree's subjective probability of a, b and c are $\frac{p}{\alpha+p+q+\alpha_{X^C}(\cdot)}$, $\frac{\alpha}{\alpha+p+q+\alpha_{X^C}(\cdot)}$ and $\frac{q}{\alpha+p+q+\alpha_{X^C}(\cdot)}$, respectively. But by Assumption 3.2, we have $\alpha_{X^C}(\cdot) = 0$.

[29] If $t < 0$, then it is equivalent to say $-t$ is the amount of transfer from the insurer to the insuree.

In this contract, the insuree gives the insurer the *net benefit* t_ω at contingency ω. Put it differently, the insurer charges the *premium* t_c to the insuree and transfers the *gross benefit* $t_c - t_a$ to the insuree when there is a fire and $t_c - t_b$ when there is a flood. The insurer's profit in expectation is therefore

$$pt_a + (1-p-q)t_b + qt_c.$$

Since the flood contingency b is announced in $C^{\{a,b,c\}}$, the insuree becomes fully aware and insuree's probability judgment of the set of contingencies is objective. Her expected utility level of $C^{\{a,b,c\}}$ is

$$pu(w_0 - t_a) + (1-p-q)u(w_0 - t_b) + qu(w_1 - t_c).$$

The outside option of the insuree is not buying the insurance, that is, $\bar{t} = 0$. If she rejects the contract, she receives her objective utility level

$$pu(w_0 - \bar{t}) + (1-p-q)u(w_0 - \bar{t}) + qu(w_1 - \bar{t}) = (1-q)u(w_0) + qu(w_1).$$

The insurer maximizes his expected profit subject to the insuree's participation constraint, that is, he solves the following problem:

$$\begin{aligned} &\max_{t_a, t_b, t_c} pt_a + (1-p-q)t_b + qt_c \qquad (3.13) \\ \text{s.t. } &pu(w_0 - t_a) + (1-p-q)u(w_0 - t_b) + qu(w_1 - t_c) \geq (1-q)u(w_0) + qu(w_1). \end{aligned}$$

The problem degenerates to a standard insurance contract in which both insurer and insuree share the same probability judgment. The solution is characterized by $w_0 - t_a = w_0 - t_b = w_1 - t_c$ together with the binding participation constraint of problem (3.13). We therefore obtain the full insurance result. Since the insuree becomes fully aware, the solution is independent of α.

Case 2: Vague Contracts

Secondly, we consider that the insurer proposes a vague contract:

$$C^{\{a,c\}} = \begin{pmatrix} a, & t_a \\ b, & t_b \\ c, & t_c \end{pmatrix} = \begin{pmatrix} \text{fire,} & t_a \\ \text{calamity but not fire,} & t_b \\ \text{no calamity,} & t_c \end{pmatrix}.$$

Although the event lists in $C^{\{a,b,c\}}$ and $C^{\{a,c\}}$ have the same reduced mapping, they are framed differently. In $C^{\{a,c\}}$, flood is expressed as “calamity but not fire” while in $C^{\{a,b,c\}}$ flood is explicitly announced. After $C^{\{a,c\}}$ is proposed, the insuree remains unaware of b. The insuree believes that buying the insurance $C^{\{a,c\}}$ leads to a utility level

$$\frac{p}{p+\alpha+q}u(w_0-t_a)+\frac{\alpha}{p+\alpha+q}u(w_0-t_b)+\frac{q}{p+\alpha+q}u(w_1-t_c).$$

If she rejects the contract, she believes that she receives her subjective utility level

$$\frac{p+\alpha}{p+\alpha+q}u(w_0)+\frac{q}{p+\alpha+q}u(w_1).$$

Thus the insurer solves the following problem:

$$\max_{t_a,t_b,t_c} pt_a+(1-p-q)t_b+qt_c \tag{3.14}$$
$$\text{s.t. } pu(w_0-t_a)+\alpha u(w_0-t_b)+qu(w_1-t_c)\geq (p+\alpha)\,u(w_0)+qu(w_1).$$

The solution is characterized by the following equation system:

$$pu(w_0-t_a)+\alpha u(w_0-t_b)+qu(w_1-t_c)-(p+\alpha)u(w_0)-qu(w_1)=0, \tag{3.15}$$
$$(1-p-q)u'(w_0-t_a)-\alpha u'(w_0-t_b)=0, \tag{3.16}$$
$$u'(w_0-t_a)-u'(w_1-t_c)=0. \tag{3.17}$$

Equation (3.15) is nothing but the binding participation constraint of problem (3.14). Equation (3.17) implies that the insuree has the same final monetary value in contingency a and c, since she puts the correct relative weights on two contingencies. However, by (3.16), if $\alpha<1-p-q$, then we have $u'(w_0-t_b)>u'(w_0-t_a)$ that implies $u(w_0-t_b)<u(w_0-t_a)=u(w_1-t_c)$. Therefore, if the insuree underestimates the existence of unforeseen calamities, she is under insured at b.

Let $\underline{\alpha}$ be a particular value of α such that equations (3.15)-(3.17) are satisfied and, additionally, $t_b=0$. Formally, $\underline{\alpha}$ satisfies

$$pu(w_0-t_a)+\underline{\alpha}u(w_0)+qu(w_1-t_c)-(p+\underline{\alpha})u(w_0)-qu(w_1)=0, \tag{3.18}$$
$$(1-p-q)u'(w_0-t_a)-\underline{\alpha}u'(w_0)=0, \tag{3.19}$$
$$u'(w_0-t_a)-u'(w_1-t_c)=0. \tag{3.20}$$

In other words, if $\alpha = \underline{\alpha}$, the insuree has zero *net* benefit at the unforeseen contingency b and thus is completely uninsured at the that contingency.[30] We now have the following lemmas.

Lemma 3.3 $\underline{\alpha} < 1 - p - q$.

Proof 3.6 *See Appendix A.1.4.*

Lemma 3.3 says, in the situation where the insuree is completely uninsured at b, the insuree must underestimates the existence of unforeseen calamities.

Lemma 3.4 $\alpha \geq \underline{\alpha}$ *if and only if* $t_b \leq 0$ *in the solution of problem (3.14).*

Proof 3.7 *See Appendix A.1.5.*

Lemma 3.4 says under the condition that the insuree's degree of AU exceeds the level in which she is completely uninsured at b, the insuree always receives a positive net benefit at b in the solution of (3.14). Furthermore, this condition is also necessary for a positive net benefit at b.

Lemma 3.5 *The insurer's profit in the solution of problem (3.14) is increasing in* α *when* $\alpha > \underline{\alpha}$ *and decreasing in* α *when* $\alpha < \underline{\alpha}$.

Proof 3.8 *See Appendix A.1.6.*

Lemma 3.5 is surprising. It implies that the insurer gains his minimal profit when $\alpha = \underline{\alpha}$ if the contract is vague. In other words, the situation where the insuree is completely uninsured at b is the worst case for the insurer.

To Be or Not to Be Vague?

After obtaining the optimal contracts in two different framings, we now examine which framing is optimal. The following proposition provides the answer.

Proposition 3.4 *There exists* $\alpha^* < \underline{\alpha}$ *such that the insurer will announce* b *in the optimal contract if and only if* $\alpha \in [\alpha^*, 1 - p - q]$.

Proof 3.9 *See Appendix A.1.7.*

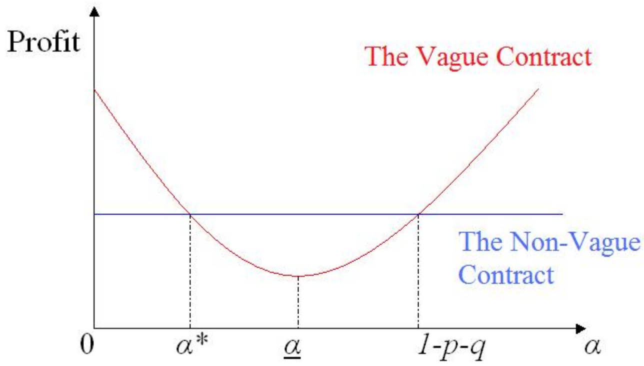


Figure 3.3: The profit curves in case 1 and 2 with different α in section 3.2.6

Figure 3.3 depicts Proposition 3.4 graphically.

For example, let $p = q = \frac{1}{3}$, $u(\cdot) = \ln(\cdot)$, $w_0 = 1$ and $w_1 = 2$. The insurer will announce b in the optimal contract if and only if $0.1596 \leq \alpha \leq \frac{1}{3}$.[31]

The following corollary directly follows proposition 3.3 and proposition 3.4.

Corollary 3.1 *There exists $\alpha^* < \underline{\alpha}$ such that the insuree cannot be exploited by the insurer if and only if $\alpha \in [\alpha^*, 1 - p - q]$.*

The interpretation of Proposition 3.4 and Corollary 3.1 is as follows.

If $\alpha > 1 - p - q$, that is, the insuree overestimates the existence the other potential calamities, "flood" does not appear in the optimal contract. Since the insuree is over-worried about the potential unknown calamities, she puts a higher weight on the other calamity event. However, in the objective world, the calamity is not so likely to occur. Thus the insurer can charge a higher premium t_c to the insuree by raising $-t_b$. The insuree is therefore over-insured at contingency b. By corollary 3.1, the insurer exploits the insuree.

However, psychological evidences suggest that $\alpha \leq 1 - p - q$. (See, e.g., Tversky and Koehler, 1994) It implies that the exploitative contract $C^{\{a,c\}}$ in case 2 when $\alpha > 1-p-q$ is not so likely to occur.

Johnson et al. (1993) provide some evidences to show that isolation of vivid causes of death increases the insuree's valuation of insurance. In this context, it means that, given the same gross benefits $t_c - t_a$ and $t_c - t_b$ in two contracts $C^{\{a,b,c\}}$ and $C^{\{a,c\}}$, the insurer

[30] Put it differently, when $\alpha = \underline{\alpha}$, the insuree pays the premium t_c to the insurer. If b occurs, then the insurer returns the premium t_c back to the insuree.

[31] Note that the result in case 1 is a special case of the result in case 2 when $\alpha = 1 - p - q = \frac{1}{3}$.

can charge a higher premium t_c to the insuree in contract $C^{\{a,b,c\}}$ where the vivid flood contingency b is announced. It is indeed the case in our model when $\alpha < 1 - p - q$.[32]

However, it does not imply that proposing $C^{\{a,b,c\}}$ is always optimal when $\alpha < 1-p-q$. It is because t_a and t_b are also endogenous variables for the insurer.

Particularly striking is that if $\alpha < \alpha^*$, that is, the insuree significantly underestimates the existence the other potential calamities, $C^{\{a,c\}}$ is better than $C^{\{a,b,c\}}$ for the insurer. The intuition is that the insuree believes that the event "calamity but not fire" is very rare. Then the insurer can provide a contract with a low gross benefit $t_c - t_b$ at flood. The insuree will accept the contract. However, her objective expected utility of the contract is lower than the objective utility level of the outside option. Thus the insurer earns a high profit by exploiting the insuree.

If $\alpha \in (\alpha^*, 1-p-q)$, that is, the insuree underestimates its existence but not too much, then "flood" appears in the optimal contract. The intuition is that α is not too low. There is no opportunity for the insurer to exploit the insuree by raising t_b. α is also not too high. There is no opportunity for the insurer to increase t_a and t_c while lowering t_b. By corollary 3.1, there is no exploitation in this contract. The insurer voluntarily does not exploit the insuree.

The main lesson is that if α is large enough (but still weakly less than the true probability $1-p-q$), the insurer will not propose a vague contract, and the insuree is not exploited. Although the insuree is unaware of the particular contingency b, because she is aware that she may be unaware of something, this makes her free from exploitation. Thus there is a value of certain degree of AU.

In Ozbay (2008) and Filiz-Ozbay (2008), the equilibrium concept requires the contract is *justifiable*, namely the contract is optimal for the insurer also from the insuree's view. Appendix A.1.8 shows that under the constraint of contractual justifiability the insurer will announce b in the optimal contract if and only if $\alpha \in (0, 1-p-q]$. Hence the role of AU is more significant: the insuree is free from exploitation whenever there is a positive degree of AU and weakly underestimates the unforeseen calamities.

3.2.7 Force majeure Clauses

We consider a situation where an employer as the principal proposes a contract to a contractor as the agent to fulfill a project. Let t, which is contractible, be the contractor's input to the project. The monetary cost of input t to the contractor is $c(t)$ where $c(\cdot)$ is a smooth, strictly increasing and strictly convex function with $c(0) = 0$, $c'(0) = 0$,

[32]The reason is that, by proposing $C^{\{a,b,c\}}$ in case 1, the insuree's subjective utility of the outside option becomes objective and is therefore lower than before. Fixing $t_c - t_a$ and $t_c - t_b$, the premium t_c can be larger by slightly lowering $-t_a$ and $-t_b$.

$c'(\infty) = \infty$. Ex post, the monetary performance of the contractor is nothing but t. The contractor has an initial wealth w. The contractor is risk averse and has a v.N.M. utility $u(\cdot)$ over money where $u(\cdot)$ is a smooth, strictly increasing and strictly concave function with $u'(0) = \infty$, $u'(\infty) = 0$. The employer who is risk neutral charges the contractor for p ex post. In the standard problem, the employer maximizes p subject to the contractor's participation, that is, he solves the following problem:

$$\max_{p,t} p$$
$$\text{s.t. } u(w + t - p - c(t)) \geq u(w).$$

Let the contractor's outside option be $\bar{p} \equiv 0$ and $\bar{t} \equiv 0$, thus her utility level of her outside option is $u(w)$ where $w > 0$ is the contractor's wealth. The solution to this problem is characterized by $c'(t^*) = 1$ and $p^* = t^* - c(t^*)$.

However, the simple situation above is in a world without *force majeure*, that is, there is no unexpected event beyond the control of contracting parties, such as war, strike, riot, crime, act of God (e.g., fire, flood, etc.). If a force majeure event occurs, the performance of the project will be jeopardized. Thus, in many contracting situations, the parties specify a force majeure clause in contract to release the contractor's obligations.

Suppose the set of contingencies is $\Omega = \{a, b\}$. The contingency a is the non force majeure contingency. For simplicity, we assume only one contingency of force majeure: b, say the contingency of "fire" in the workplace of the project. The general event "force majeure" is $X = \{b\}$. Let $\mu(\{a\}) = q$ and $\mu(\{b\}) = 1 - q$.

Before contracting, the contractor is fully aware of contingencies a while she is unaware of contingency b, that is, $K_0 = \{a\}$. But she is aware that there may be some unforeseen force majeure contingencies.

Again, we have $\alpha_X(K) = 0$ for $b \in K$, that is, if the fire contingency b is announced, the contractor will be fully aware of b, and then she comprehends the objective probability space (Ω, μ).

In contrast, $\alpha_X(K) \equiv \alpha$ for $b \notin K$. It captures the degree that the contractor is aware that she may be unaware of some particular force majeure contingencies if b is not announced. The contractor assigns a weight α to the unforeseen event force majeure $\{b\}$ while assigning weight q to contingency a. Thus her subjective probability of b is $\frac{\alpha}{\alpha+q}$, and her subjective probability of a is $\frac{q}{\alpha+q}$.

If a force majeure occurs, the contractor's performance in terms of money is zero, that is, the contractor's performance is totally destroyed in force majeure events. If there is no force majeure, the monetary performance equals to the input t.

The contractor's v.N.M utility function is therefore

$$u_A(\omega, t) \equiv \begin{cases} u(w - p - c(t)) & \text{for } \omega \in X \\ u(w + t - p - c(t)) & \text{for } \omega \in X^C \end{cases}.$$

The employer's v.N.M utility is $u_P(p, t) = p$.

The contractor implements the input t before the contingency is revealed. Thus t is forced to be identical in all contingencies. Formally, the set of admissible contracts with K is $\mathbb{C}^K = \{C^K : C^K(a) = (p_1, t) \text{ and } C^K(b) = (p_2, t)\}$. Note that Assumption 3.3 is satisfied here.

We now consider that the employer proposes a vague contract $C^{\{a\}}$ where b is not announced but a force majeure clause is specified:

$$C^{\{a\}} = \begin{pmatrix} a, & (p_1, t) \\ b, & (p_2, t) \end{pmatrix} = \begin{pmatrix} \text{not force majeure,} & (p_1, t) \\ \text{force majeure,} & (p_2, t) \end{pmatrix}.$$

Facing $C^{\{a\}}$, the contractor is still unaware of b. Since her outside option is $\bar{p} \equiv \bar{t} \equiv 0$, she believes that accepting $C^{\{a\}}$ leads to a utility level

$$\frac{q}{q+\alpha} u(w + t - p - c(t)) + \frac{\alpha}{q+\alpha} u(w - p - c(t))$$

and rejecting $C^{\{a\}}$ leads to a utility level

$$\frac{q}{q+\alpha} u(w + \bar{t} - \bar{p} - c(\bar{t})) + \frac{\alpha}{q+\alpha} u(w - \bar{p} - c(\bar{t})) = u(w).$$

Thus the employer solves the following problem:

$$\begin{aligned} &\max_{p_1, p_2, t} qp_1 + (1-q)p_2 \\ \text{s.t. } &\frac{q}{q+\alpha} u(w + t - p_1 - c(t)) + \frac{\alpha}{q+\alpha} u(w - p_2 - c(t)) \geq u(w). \end{aligned} \tag{3.21}$$

It is straightforward to show that the solution of problem (3.21) is p_1^*, p_2^* and t^*, which are characterized by the following equation system:

$$c'(t^*) - q = 0,$$

$$(1-q)u'(w + t^* - p_1^* - c(t^*)) - \alpha u'(w - p_2^* - c(t^*)) = 0, \tag{3.22}$$

$$\frac{q}{q+\alpha} u(w + t^* - p_1^* - c(t^*)) + \frac{\alpha}{q+\alpha} u(w - p_2^* - c(t^*)) - u(w) = 0. \tag{3.23}$$

Observation 3.1 *If $\alpha < 1 - q$, then $u(w + t^* - p_1^* - c(t^*)) > u(w - p_2^* - c(t^*))$.*

We obtain observation 3.1 from equation (3.22). If $\alpha < 1-q$, then $u'(w+t^*-p_1^*-c(t^*)) < u'(w-p_2^*-c(t^*))$. Thus we have $u(w+t^*-p_1^*-c(t^*)) > u(w-p_2^*-c(t^*))$. The condition $\alpha < 1 - q$ that is suggested by the psychology literature (See Tversky and Koehler, 1994) means that the contractor underestimates the existence of the force majeure. $u(w + t^* - p_1^* - c(t^*)) > u(w - p_2^* - c(t^*))$ is a result where the contractor is not fully "insured". The contractor is better off in the non force majeure event that is more likely in reality. Hence, besides hidden information and hidden action, underestimating unforeseen contingencies can be also a driving force for a non-full insurance outcome.

However, if b is announced in the contract, then the contractor shares the same probability measure as the employer. It is then equivalent for the employer to solve problem (3.21) when $\alpha = 1-q$. The question is when the employer has an incentive to make the contractor have the correct belief. The following proposition provides a negative answer.

Proposition 3.5 *If $\alpha \neq 1 - q$, then proposing a vague contract $C^{\{a\}}$ is always better than a non-vague contract $C^{\{a,b\}}$ for the employer.*

Proof 3.10 *See Appendix A.1.9.*

Proposition 3.5 says that if the contractor is unaware of b, no matter how aware of her unawareness she is, the employer will be silent on the particular force majeure contingency fire and describes only a general force majeure event in the contract. Since the contract is always vague, by proposition 3.3, the contractor is always exploited. If $\alpha < 1 - q$, that is, the contractor underestimates the existence of force majeure, the employer exploits the contractor by charging a high p_2, which occurs more likely than the contractor believes. If $\alpha > 1 - q$, the employer exploits the contractor by charging a high p_1, which also occurs more likely than the contractor believes. Thus whenever the contractor is unaware of the particular force majeure contingencies, the employer can utilize the contractor's mis-perception.

For example, let $u(\cdot) = \ln(\cdot)$, $c(t) = \frac{1}{2}t^2$, $q = \frac{1}{2}$ and $w = 1$. Then the profit function of the employer which depends on α is depicted in Figure 3.4. Note that the result when b is announced is a special case of this result when $\alpha = 1 - q = \frac{1}{2}$. In Figure 3.4, we observe that two profit curves intersect at $\alpha = \frac{1}{2}$. Moreover, we find that the employer's profit when announcing b is always higher than not announcing it, and two curves are tangent at $\alpha = \frac{1}{2}$.

Proposition 3.5 presents a negative result. The only way to make the contractor free from exploitation is making the contractor aware in order to let her have a correct probability judgment. In contrast to the result in the insurance example, some certain degree of the

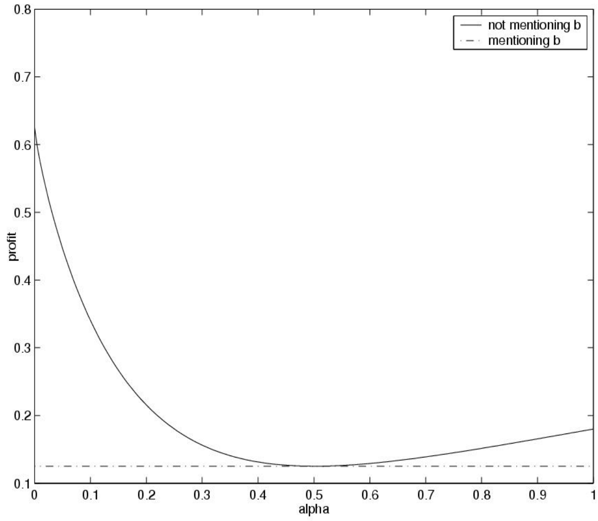


Figure 3.4: The profit curves when announcing b or not with different α in section 3.2.7

contractor's AU cannot creates the incentive for the employer to make the contractor aware. Only ex ante full awareness of the contractor is valuable to her.

3.2.8 Persuasive Advertising

In this section, we consider a firm as the principal sells an experience good to a consumer as the agent.[33] Suppose now the firm provides a travel service. The set of contingencies is $\Omega = \{g, b\}$. The contingency g is the contingency of the consumer's good experience. b is the contingency of her bad experience. For simplicity, let announcing g in our language be a full description of the good contingency of travel, say an *advertisement* showing the most beautiful sites with sunshine. On the other hand, announcing b is the full description of the bad contingency, say expressing the possibility of a storm, theft and so on. The general event is a "good" experience $X = \{g\}$. Let the probability measure be $\mu(\{g\}) = q$ and $\mu(\{b\}) = 1 - q$.

The reason we focus on the experience good here is that, before contracting, the consumer knows nothing about the content of the travel. She is aware of no contingencies, that is, $K_0 = \emptyset$. In reality, travelers enjoy mainly the unknown experiences during the travel. A contingent plan under uncertainty would be uninteresting for the travelers. But the consumer has the general idea of events X and X^C, that is, a good experience and a bad experience.

[33]The experience good is a product or service whose payoff-relevant characteristics are difficult to know in advance. The typical examples are travel, movie, etc.

By Assumption 3.2, we have that $\alpha_X(K) = 0$ for $g \in K$ and $\alpha_{X^C}(K) = 0$ for $b \in K$. If g (respectively b) is explicitly described in the contract, the consumer will be fully aware of g (respectively b), and assigns a correct weight $\mu(g) = q$ to g (respectively $\mu(b) = 1 - q$ to b) and zero weight to the non-existing residual event. We call the contract $C^{\{g\}}$ describing the contingency g a *positive* advertisement and $C^{\{b\}}$ a *negative* advertisement. Suppose for simplicity that the cost of advertisement is zero.

Let $\alpha_X(K) \equiv \alpha_g$ for $g \notin K$ ($\alpha_{X^C}(K) \equiv \alpha_b$ for $b \notin K$). Assume that α_g, $\alpha_b > 0$. It captures the fact that, if the contract is vague in the good experience X (respectively bad experience X^C), the consumer is aware that she may be unaware of some particular good contingencies (respectively bad contingencies).

The choice of the firm is a pair (p, t) where p is the monetary transfer from the consumer to the firm, or the price of travel and $t \in \{0, 1\}$ is the consumer's binary choice of accepting the travel or not. The cost of the travel is zero. The firm's utility is $u_P(p, t) = pt$. If $t = 1$, that is, the consumer accepts the contract, the firm receives the price p. Otherwise, he gets zero. Let $v > 0$ denote the consumer's valuation of the good experience. Assume the consumer's valuation of the bad experience is zero.

The consumer's v.N.M. utility function is therefore

$$u_A(\omega, p, t) \equiv \begin{cases} t(v - p) & \text{for } \omega = g \\ t(0 - p) & \text{otherwise} \end{cases}.$$

If the consumer accepts the contract ($t = 1$), the consumer's utility is her benefit from the travel v net of the price p when a good experience occurs. When a bad experience occurs, the consumer pays the price p but gains nothing.

The consumer has to decide whether or not to accept the contract before the contingency is revealed. Thus t and p must be identical in all contingencies. Formally, the set of admissible contracts with K is $\mathbb{C}^K = \{C^K : C^K(\omega_1) = C^K(\omega_2) \text{ for all } \omega_1 \neq \omega_2\}$. (Assumption 3.3 is satisfied.)

- Vagueness in both Good and Bad Experiences:

Firstly, we consider a case in which the firm does not advertise, that is, neither g nor b is announced. The contract is $C^{\emptyset}(\cdot) = (p, t)$. There is only one "catchall" clause in $C^{\emptyset}$ that is what to do no matter what happens.

The consumer believes that accepting the contract leads to a utility level

$$\frac{\alpha_g}{\alpha_g + \alpha_b} t(v - p) + \frac{\alpha_b}{\alpha_g + \alpha_b} t(-p).$$

On the other hand, let the consumer's outside option be $\bar{p} \equiv 0$ and $\bar{t} \equiv 0$. If she rejects the contract, she believes that she receives her utility level

$$\frac{\alpha_g}{\alpha_g + \alpha_b}\bar{t}(v - \bar{p}) + \frac{\alpha_b}{\alpha_g + \alpha_b}\bar{t}(-\bar{p}) = 0.$$

The firm therefore simply solves the following problem:

$$\max_{p,t} pt$$
$$\text{s.t. } \frac{\alpha_g}{\alpha_g + \alpha_b}t(v - p) + \frac{\alpha_b}{\alpha_g + \alpha_b}t(-p) \geq 0.$$

The solution to this problem is $p_1 = \frac{\alpha_g}{\alpha_g+\alpha_b}v$ and $t_1 = 1$. The firm charges the consumer the price $\frac{\alpha_g}{\alpha_g+\alpha_b}v$ and the consumer accepts it. The corresponding profit for the firm is $\pi_1 = p_1 = \frac{\alpha_g}{\alpha_g+\alpha_b}v$.

- Vagueness in Bad Experiences:

Secondly, we consider that the firm makes only the positive advertisement, that is, only g is announced. The contract is $C^{\{g\}}(\cdot) = (p, t)$, which frames the good experience differently. After understanding $C^{\{g\}}$, the consumer has a correct weight to contingency g. Then the firm solves the following problem:

$$\max_{p,t} pt$$
$$\text{s.t. } \frac{q}{q + \alpha_b}t(v - p) + \frac{\alpha_b}{q + \alpha_b}t(-p) \geq 0.$$

The solution to this problem is $p_2 = \frac{q}{q+\alpha_b}v$ and $t_2 = 1$. The corresponding profit of the firm is $\pi_2 = \frac{q}{q+\alpha_b}v$.

- Vagueness in Good Experiences:

Similarly, if the firm makes only the negative advertisement, the firm charges price $p_3 = \frac{\alpha_g}{\alpha_g+1-q}v$ and $t_3 = 1$. The corresponding profit of the firm is $\pi_3 = \frac{\alpha_g}{\alpha_g+1-q}v$.

- No Vagueness:

Lastly, if both g and b are announced, the contract $C^{\{g,b\}}(\cdot) = (p, t)$ is not vague. The consumer then understands the probability space. The solution for the firm is $p_4 = qv$ and $t_4 = 1$. The corresponding profit of the firm is $\pi_4 = qv$.

Proposition 3.6 *The contract $C^{\{g\}}$ is optimal for the firm if and only if $\alpha_g < q$ and $\alpha_b < 1 - q$.*

Proof 3.11 *See Appendix A.1.10.*

Proposition 3.6 says that if the consumer underestimates both positive and negative concrete scenarios of the good, then it is optimal for the firm to make only the positive advertisement. The intuition is straightforward. By making only the positive advertisement, the consumer puts a high weight on the good contingency g. The firm can therefore charge the highest price. Conversely, if it is optimal for the firm to make only the positive advertisement, the consumer necessarily underestimates both positive and negative contingencies of the good. Since we observe that in most advertisements only the good contingencies are announced in reality, we also confirm consumers' psychological characteristic that they underestimate both contingencies.

Consequently, the consumer's *subjective* valuation of the good is higher. This is a typical *persuasive advertising* result. However, the welfare implication is that the persuasive advertising on experience good is harmful to the consumers. Since $C^{\{g\}}$ is vague in $\{b\}$, by proposition 3.3, the consumer is exploited. In the result, $p = \frac{q}{q+\alpha_b}v$. The *objective* expected utility level of the consumer is $qv - p = \frac{qv}{q+\alpha_b}[\alpha_b - (1-q)] < 0$ since $\alpha_b < 1 - q$. The objective participation constraint of the consumer is violated. Thus such persuasive advertising hurts the consumers. In the standard persuasive advertising literature, it is difficult to judge the consumer's welfare change after persuasive advertising, since it is not clear we should use the utility before advertising or after it as the welfare criterion. This simple example suggests that neither of them should be the criterion since they are both subjective, but there exists an objective one known only by the firm and the fully aware consumers.

Hence, the policy recommendation is that, if competition among firms is not possible, the firm is required to report the bad contingencies compulsorily in the advertisement. For instance, it has been already mandatory to include a health warning in the Tobacco advertising in many countries.

However, if we extend the model by introducing Bertrand-competition among homogeneous firms, in equilibrium, all firms will choose $p = 0$. For every firm, each framing is possible to occur in equilibrium. The vagueness of the contract changes the consumer's ex ante subjective utility of contracts but plays no role in competition. Since $p = 0$, the consumer's objective expected utility is maximized irrespective of her ex ante subjective valuation of the good. Thus competition does not necessarily promote awareness of the consumer, but increases the consumer's welfare to the best extent.

3.2.9 Framing the Future and Self-Control Problems

In this section, we consider a benevolent principal encourages a present-biased agent to perform a long-run goal. For example, some parents want to stimulate their kid to study harder, and someone may want to encourage his friend to achieve an ambitious task. There is no conflict of interest between the principal and the agent. The principal's motivation of manipulating the agent's belief here is in order to help the agent overcome her self-control problem.

The set of contingencies is $\Omega = \{g, b\}$. The contingency g is the good contingency: the task is successful. b is the failure contingency. If the agent exerts efforts, the principal's utility (or the agent's objective utility) is $pv - 1$ where $v > 1$ is the benefit of the task, 1 is the cost of efforts and p is the objective probability of success. Let $K^0 = \emptyset$, that is, the agent initially knows nothing about the future. But she has a general event $X = \{g\}$ in mind. The agent knows that after making the effort something good or bad will occur in the future. Let $\alpha_X(K) \equiv \alpha_g$ for $g \notin K$ ($\alpha_{X^C}(K) \equiv \alpha_b$ for $b \notin K$). Assume that α_g, $\alpha_b > 0$ and $\alpha_g < p$, $\alpha_b < 1-p$ as usual. Before contracting, the agent's subjective utility is $\beta\frac{\alpha_g v}{\alpha_g+\alpha_b} - 1$ where $\beta < 1$ represents the present bias of the agent.

Suppose $pv - 1 > 0$ and $\beta\frac{\alpha_g v}{\alpha_g+\alpha_b} - 1 < 0$. Thus the agent "should" exert efforts, but she is too lazy to do it. To make the thing interesting, we assume further that $\beta pv - 1 < 0$. That is, if the agent is fully educated about the future, she still prefers not performing the task because of her present bias. However, the following proposition provides a solution to the agent's self-control problem.

Proposition 3.7 *If $p(\beta v - 1) > \alpha_b$, only the contract $C^{\{g\}}$ overcomes the agent's self-control problem.*

Proof 3.12 *Straightforward and omitted.*

Proposition 3.7 says that if the self-control problem of the agent is not so severe ($\beta v - 1 > 0$ and it is large enough) and the both contingencies are substantially underestimated (p is actually large and α_b is small), then the principal can describe only the good scenario and shroud the bad scenario so as to motivate the agent. This manipulation of the agent's belief makes the agent more optimistic about the future, since the agent's subjective probability of success is $\frac{p}{p+\alpha_b} > p$. But this mis-perception can overcome the agent's self-control problem. A similar idea is also in Benabou and Tirole (2002) where overconfidence of one's ability is valuable.

3.2.10 Contracts and Interpretation

Interpretation has been the subject in linguistics, philosophy of language, theology, psychology and legal doctrine for a long time. In legal literatures, the issue of *contract interpretation* is frequently discussed.[34] In economics, *Contract theory* is one of the most active and successful research fields.[35] Unfortunately, issue of contract interpretation is almost neglected in economic literature.

Parallelly, there is a growing literature in *language and economics*.[36] Many economists realize that natural language plays an important role in the process of decision making and transaction. It seems that a "linguistic turn" in economics occurs.[37] As Lipman states that "The world people live in is a world of words, not functions, $\cdots\ \cdots$, real contracts are written in a language." (See Rubinstein (2000) pp. 114) Nevertheless, there are so few papers in contract theory where natural language plays a role. Thus it is indeed too early for me to write a survey in a field that just starts.

Related Literature:

Though the paper focuses on economic analysis of contract interpretation, there are some other related literatures in incomplete contracting worthy to be mentioned.

Anderlini and Felli (1999) and Battigalli and Maggi (2002) are among the papers that endogenize contractual incompleteness based on explicit contracting costs.

Anderlini and Felli (1999) model contractual incompleteness endogenously from the complexity costs. There are ex ante writing costs and ex post implementing costs. They consider a simple bilateral risk-sharing model. The optimal sharing rule is incomplete in the sense that it does not prescribe the best allocations for some states of nature.

Battigalli and Maggi (2002) model contractual incompleteness endogenously from the costs of writing contracts. Due to the costs of describing the events and the parties' behavior, the model predicts two kinds of incompleteness: *discretion*, meaning that the behaviors of parties are not specified in details; and *rigidity*, meaning that the obligations of parties are not contingent on the detailed state of nature. More interestingly, they use syntactic approach to model events and parties' behavior in which the language written in contracts consists of primitive sentence and logical connectives. It is the only paper I am aware of , in which the written contract is very close to our natural language in the field of contract theory.

[34]See Schwartz and Scott (2003), Posner (2004) and the references in these papers.

[35]See a comprehensive textbook by Bolton and Dewatripont (2005) for reference.

[36]See the book by Rubinstein (2000) and the survey by Lipman (2003) for reference.

[37]There was a "linguistic turn" in philosophy and many other areas in social science in the beginning of 20th century.

However, these papers assume that the "language" in contracts is common-language for both parties and court. The court is *passive* enforcer of contracts. Thus there is no room for contract interpretation of the court.

Anderlini, Felli and Postlewaite (2006) is one of the few papers in which the court plays an *active* role. They analyze a contractual environment where voiding some contract by the court improves the ex-ante welfare of contracting parties. They show that if the court simply enforces all contracts and a pair of contracting parties are asymmetrically informed at the contracting stage, then there is pooling equilibrium. But if the court voids some voluntary contract, a separating equilibrium emerges. Although such voidance leads to a welfare loss, the *net ex ante welfare* is enhanced. However, this paper only considers the court's role of voiding some widget. There is no room for contract interpretation in general.

Of course, there are many other important papers related to this survey. Due to the limit of my knowledge, I could not cover all of the them. In the short survey, I will introduce formally only two works on economics of contract interpretation: Shavell (2006) and Heller and Spiegler (2008). Shavell (2006) is the first paper providing a general analysis of the role of courts in contract interpretation through an economic model. The idea is that the court chooses the optimal interpretation method to maximize the social welfare, given the parties' best contracting response to the method of interpretation. The model predicts that, in the presence of writing costs, the contracting parties optimally write incomplete contracts that include gaps and fairly general terms. The court should not always enforce what parties write in contracts. Heller and Spiegler (2008) endogenize a special kind of contractual incompleteness: *contradiction*, given the special method of interpretation: *precedent system*. In these two models, courts play a very important role in contract interpretation.

The paper is structured as follows. Section 3.2.10 argues why the issue of contract interpretation is relevant to contracting problem. Section 3.2.10 presents an economic model by Shavell (2006) which provides a general analysis of contract interpretation. In section 3.2.10, we answer the question why there are contradictions in contracts based on Heller and Spiegler (2008). The last section concludes.

Why does Contract Interpretation matter?

Schwartz and Scott (2003) provides a normative theory of contract law. They argue that the meta principle of contract law is to maximize the joint surplus of contracting parties from transactions.

The first order implication of the goal is the perfect enforcement of all *efficient* transactions by the state. Standard contract theory presumes a perfect enforcement of all kinds of contracts. However, if the principal normative claim is to maximize contracting parties' welfare, the perfect enforcement might not be recommended in some inefficient transactions. In section 3.2.10, we could see the justification for the correction of some written contracts by courts.

Moreover, the court could not enforce the contract without understanding it. For example, if there are gaps and contradictions in the written contract, it is vacant for the court to enforce it. The court needs to define a method of interpretation that maps from written contracts to their legal implications. The court needs to interpret the contract before enforcing it. However, a second order implication of primary goal, "interpretation", is missing in standard contract theory.

How Should We Write and Interpret Contracts?

Why do parties write gaps and fairly general terms in contracts? How should the court interpret them? Should courts always enforce what contracting parties write in contracts? In the following subsection, we answer these questions based on Shavell (2006).

A Model Throughout the paper, assume that there are two agents involved: an *interpreter* (or a court) and a *writer* (or a pair of contracting parties). The interaction of the writer and the interpreter is modeled by a two-stage situation. The interpreter chooses the interpretation method firstly and then, based on the interpretation method, the writer writes a contract.

The set of *states of nature* is a discrete set Ω. We denote its element by ω_i with $i \in \{1, \cdots, n\}$. So there are totally n states. The probability of ω_i is p_i. A writer is identified by its *type* t with $t \in T$. The probability of type t is $q_t > 0$. t is not observable by the interpreter. We denote the *action* of a writer by a with $a \in A$. The *utility* of the writer or joint surplus of the contracting parties is $u(a, \omega_i, t)$ if her type is t, her action is a and the state is ω_i. The action that maximize $u(a, \omega_i, t)$ for given ω_i and t is called *ideal action* $a^*(\omega_i, t)$.

A *contract* is characterized by $C = \{c_1, \cdots, c_K\}$ where $c_k = (e_k, a_k)$ with $e_k \subseteq \Omega$. In words, the contract is a list of clauses. A *clause* c_k is a pair of *event* e_k and the action a_k to be taken when e_k occurs. e_k is also called a *term*. We call a term e_k a *specific* term if e_k contains a single state ω_i, otherwise the term is called *general* term. Let $E^C = \{e_1, \cdots, e_K\}$ denote the collection of all terms in contract C. For every $\omega_i \in \Omega$, let $J^C(\omega_i) = \{k \in \{1, \cdots, K\}\colon\ \omega_i \in e_k\}$. In words, $J^C(\omega_i)$ is the set of terms that cover ω_i in contract C. We assume that $|J^C(\omega_i)| \leq 1$ for all ω_i and C in this section. Thus all terms e_k in any contract C are mutually exclusive events. Put it differently, contradictory clauses are impossible. However, all terms e_k in a contract C are not necessarily exhaustive. If there exists some ω_i s.t. $|J^C(\omega_i)| = 0$, ω_i falls into a *gap*, because ω_i is not covered by any clause. We assume that there is a *writing cost* α per term. The number of terms in a contract C is $z(C)$. Thus the total writing cost of the writer is $\alpha z(C)$.

The *method of interpretation* is characterized by a function I which maps a *written contract* C to a *interpreted contract* $I(C)$ where $I(C) = I(\{(e_1, a_1), \cdots, (e_K, a_K)\}) = \{(\omega_1, a(\omega_1)), \cdots, (\omega_n, a(\omega_n))\}$. In words, the interpreter interprets any written contract to a complete contract which assigns an unique action to each state of nature. Since t is not observable by the interpreter, $a(\cdot)$ in $I(C)$ is independent of t. It is assumed that the interpretation method displays *independence* which says that, for all ω_i, $a(\omega_i)$ is independent of contractual terms that do not cover ω_i and that how a gap is filled is independent of contractual terms. Thus it excludes the possibility that the writers could *signal* their types. Furthermore, it is assumed that ex post renegotiation is not allowed.

Given the interpretation method $I(\cdot)$, the type t writer's gross *expected utility* of a written contract C is $U(I(C), t) = \sum_\Omega p_i u(a(\omega_i), \omega_i, t)$. The net expected utility is $U(I(C), t) - \alpha z(C)$. Thus given $I(\cdot)$, the type t writer chooses C to maximize her net expected utility. We denote the optimal written contract by $C(I, t)$. The value function of the writer is $V(I, t) = max_C U(I(C), t) - \alpha z(C)$. We assume that the interpreter is a benevolent policy maker. She chooses $I(\cdot)$ to maximize the social welfare $W(I) = \sum_T q_t V(I, t)$.

Theorem 3.2 *Literal enforcement of contracts as written is not optimal.*

Proof 3.13 *Let the interpretation method $I(\cdot)$ be literal enforcement of contracts as written. $I(\cdot)$ implies that there is no gap in the contract. Suppose, for some type t, a writer has a clause (e_k, a_k) in her optimal written contract $C(I, t)$. She will get the utility level $V(I, t)$. Consider another interpretation method $\hat{I}(\cdot)$ in which all terms are interpreted as written and fill gaps for all ω_i in e_k with action a_k. Thus, given $\hat{I}(\cdot)$, type t writer could be strictly better off if she writes the same contract as before except leaving gaps for all ω_i in e_k because she saves one term writing cost. We have $V(\hat{I}, t) \geq V(I, t) + \alpha$ for type t writer. Moreover, all other types will be at least as well off as before, as they can write the same contract as before. Thus the social welfare $W(\hat{I}) > W(I)$. $I(\cdot)$ is not optimal.*

Now a natural question arises: what is the optimal way to fill gaps? For simplicity, suppose a fraction f of writers want action a_1 in a state ω_i and would get utility u_1 from a_1 and 0 from the other action a_2. The remaining fraction $1 - f$ of writers want action a_2 in ω_i and would get utility u_2 from a_2 and 0 from a_1. It is easy to show that a gap should be filled with a_1 if and only if $f\ \min\{p_i u_1, \alpha\} > (1-f)\ \min\{p_i u_2, \alpha\}$. In words, a gap should be filled with a_1 if and only if the expected cost of filling the gap with a_1 exceeds the expected cost of filling the gap with a_2.

Theorem 3.3 *Specific terms are interpreted as written under the optimal independent method of interpretation.*

Proof 3.14 *Suppose $I(\cdot)$ is the optimal independent method of interpretation and $I(\cdot)$ overrides some specific term to be (ω_i, a). Consider another method of interpretation $\hat{I}(\cdot)$ which is the same as $I(\cdot)$, except that under $\hat{I}(\cdot)$ specific terms are interpreted as written. Now the type t writer could replace $C(I,t)$ by a new contract C which is the same as $C(I,t)$ except that the specific term is replaced by $(\omega_i, a^*(\omega_i, t))$. Since $z(C)$=$z(C(I,t))$, each writer involves the same writing cost as before. But each writer is weakly better off in state ω_i. Therefore, $I(\cdot)$ is weakly dominated by $\hat{I}(\cdot)$. $I(\cdot)$ is not optimal independent method of interpretation.*

Shavell (2006) shows that, in general, a specific term might not be interpreted as written under the optimal method of interpretation if *independence* assumption is relaxed.

Moreover, Shavell (2006) shows that under the optimal method of interpretation it is possible that a general term is interpreted as written but this is not ideal for the writer, a general term is overridden in a state but interpreting the term as written would be better for the writer. However, an *opt-out rule*, under which the writer can specify any term that they write not be overridden, is socially desirable.

Evidence Beyond the Contract Suppose now the interpreter considers not only the contract, but also the extrinsic *evidence* beyond the contract. The evidence provided by the writer could certify the writer's type t. $\beta > 0$ represents the cost of presenting the evidence. If evidence is presented, the ideal action $a^*(\omega_i, t)$ will be chosen by the interpreter. Assume further that specific terms are interpreted as written.

Now we examine how writers contemplate whether writing specific term for a state ex ante or providing the evidence in this state ex post. The evidence will not be presented for state ω_i if $p_i\beta > \alpha$, since the expected cost of presenting evidence exceeds the writing cost while the benefits of two alternatives are identical. Furthermore, evidence will be presented if and only if $\beta < u(a^*(\omega_i, t), \omega_i, t) - u(a(\omega_i), \omega_i, t)$, that is, the cost of providing evidence is less than the extra utility of doing so.

Comments A natural question is whether extrinsic evidence beyond the contract is socially desirable? Traditionally, there is a debate between two conflicting arguments in philosophy of law and legal theory. On one hand, Willistonian, or "textualist", theory of interpretation argues that the court should use only the contract to determine the plain meaning of the contracting parties. On the other hand, "contextualist" theory of interpretation argues that the court should use the broad evidentiary bases, including the contract itself and the extrinsic evidence to determine the meaning of the contract.[38]

Based on Shavell (2006), we could do a social welfare comparison of these two approaches. Then we could conclude when extrinsic evidence beyond the contract should be permitted based on this economic analysis.

Why are there Contradictions in Contracts?

In last section, we assume that $|J^C(\omega_i)| \leq 1$ for all ω_i and C. But, sometimes, a state of nature is covered by more than one terms with different actions. For example, if $C = \{(earthquake, a_1), (fire, a_2)\}$ and $a_1 \neq a_2$, then there are overlapping of terms in C. When the state is both earthquake and fire, it is ambiguous that which of a_1 and a_2 should be implemented. Heller and Spiegler (2008) allows that $|J^C(\omega_i)| > 1$. A state ω_i could be covered by multiple clauses. If the actions specified by these clauses are different, the clauses are *contradictory*.

The Model In this subsection, we will see a rationale for contradictions in contracts. We use the same notations in section 3.2.10, the difference of the model will be introduced as below.

Heller and Spiegler (2008) construct a sequential-move game. The writer chooses the contract C firstly and then the interpreter interprets it. But some restrictions are put on the interpreter's choice exogenously.

In contrast with Shavell (2006), the set of states of nature is a continuum $\Omega = [0, 1)$. A state $\omega \in \Omega$ is drawn from the uniform probability measure μ over Ω. Let $\delta(C)$ be the μ-measure of states ω for which $|J^C(\omega)| \neq 1$. In words, $\delta(C)$ measures the *degree of vagueness* of contract C. The action set A is the set of real numbers $\Re$. Now the writer is homogeneous. So we ignore variable t in this model. Moreover, the goal of interpreter is not to maximize the social welfare. There is a conflict of interests between the writer and the interpreter. The utility of j, $j \in \{writer, interpreter\}$, is $u_j(a, \omega) = -(\omega - m_j + a)^2$. Assume that $m_{writer} = 0$ and $m_{interpreter} = m$ with $m \neq 0$.

It is assumed that the linguistic resource is limited. This feature is captured by the assumption that, for all k, $e_k \in \Im$ where $\Im$ is a family of *admissible events*. $\Im$ is introduced to capture bounds on the complexity of language for describing events. Here we consider only the simplest possible family. Assume that every event in $\Im$ is a interval

[38] See Schwartz and Scott (2003).

$[x, y)$ with $x \geq 0$ and $y \leq 1$.[39] It reflects that the writer can only describe interval events. Moreover we replace the role of writing cost by a fixed number of terms K for all contracts.

Before introducing the interpretation method, we define an important concept. For all contract C and $J \subseteq \{1, \cdots, K\}$, define $s_J^C = \{\omega \in \Omega$: $\forall k \in J$ $\omega \in e_k$ and $\forall j \notin J$ $\omega \notin e_j\}$. Note that $\{s_J^C\}_{J \subseteq \{1, \cdots, K\}}$ is a partition of Ω. We call it the *effective partition* induced by E^C and denote it by S^C. A particular example when $K = 3$ is shown in Figure 3.2.10. We could see that $S^C = \{s_{\{1\}}^C, s_{\{1,2\}}^C, s_{\{2\}}^C, s_{\{2,3\}}^C, s_{\{3\}}^C, s_{\emptyset}^C\}$ is a partition of Ω. $s_{\{1,2\}}^C$ and $s_{\{2,3\}}^C$ are the contradictory events. $s_{\emptyset}^C$ is the gap area.

Give the contract C written by the writer, the interpreter chooses a function $I : S^C \mapsto A$. The property that I maps from S^C to A reflects a *precedent system*. In words, it requires the interpreter to make the same decision in two states in the same contradictory situation or in the gap. Now we put some restriction on $I(\cdot)$. Assume that, for every state ω, if $J^C(\omega) \neq \emptyset$ and $a_k = a$ for every $k \in J^C(\omega)$, then $I(s_J^C) = a$. In other words, the interpreter is allowed to exercise discretion on the action choice only if the state ω falls into a gap or contradiction.

Now we analyze the subgame perfect equilibrium of the game. Let the equilibrium contract be $C^* = \{(e_1^*, a_1^*), \cdots, (e_K^*, a_K^*)\}$. Heller and Spiegler (2008) show that each $s_J^{C^*}$ in S^{C^*} is an interval. Moreover, for any interval $[x, y) \subseteq [0, 1)$, let $a_j([x, y)) \equiv \text{argmax}_a$ $E_{[x,y)}[u_j(a, \omega)]$. Then $a_j([x, y)) = -(\frac{x+y}{2}) + m_j$. If the interpreter has discretion over the event $[x, y)$, she will choose action $-(\frac{x+y}{2}) + m$. Otherwise, the writer will choose action $-(\frac{x+y}{2})$. Thus for every $s_J^C = [x, y)$ for which $|J| \neq 1$, $I(s_J^C) = -(\frac{x+y}{2}) + m$. Otherwise, $I(s_J^C) = -(\frac{x+y}{2})$. In Figure 3.2.10, we see that if the state ω falls into $s_{\{1,2\}}^C$, $s_{\{2,3\}}^C$ or $s_{\emptyset}^C$, the interpretation is delegated to the interpreter.

If the writer creates contradictions, on one hand, she will lose from delegation to the interpreter; On the other hand, she could reduce the variance of some other element s_J^C in S^C for which $|J| = 1$. Hence the writer faces a trade-off. Since the writer is risk averse, in equilibrium, all cells $s_J^{C^*}$ with $|J| = 1$ have the same measure and all cells $s_J^{C^*}$ with $|J| \neq 1$ have the same measure as well.

We say that a contract C induces a *fuzzy partition* if there exists a numbering of the elements in E^C , such that: (i) the only allowed intersections between e_k and e_j are as follows: k and j must be consecutive, and neither interval contains the other; (ii) the gap is either $[0, x)$ or $[x, 1)$. In a fuzzy partition E^C, adjacent terms are “almost” disjoint, except that the border between them is blurred. Thus the effective partition S^C induced by E^C consists of $2K$ intervals, of which K intervals are delegated to the interpreter. For example, in Figure 3.2.10, $E^{C^*} = \{e_1^*, e_2^*, e_3^*\}$ is a fuzzy partition. The effective partition

[39]Under this assumption, there is no specific terms in contracts. Actually any specific term has probability measure zero under the continuum of states.

S^{C^*} induced by E^{C^*} has 3 delegated intervals and 3 undelegated intervals.

Heller and Spiegler (2008) show the following theorem:

Theorem 3.4 *The equilibrium contract C^* induces a fuzzy partition. All undelegated intervals in S^{C^*} are of equal measure $(1-\delta(C^*))/K$, and all delegated intervals in S^{C^*} are of equal measure $\delta(C^*)/K$ with $\delta(C^*) = max\,\{\frac{1}{2} - 2K^2m^2, 0\}$.*

In Figure 3.2.10, we see an equilibrium contract C^* with a positive degree of vagueness. In general, as K decreases (or the writing cost is higher) and as m decreases (or the conflict of interests between the writer and interpreter is lower), the equilibrium degree of vagueness is higher. Therefore, under the precedent system, if there are high writing cost and low conflict of interest, contradictions in contracts emerge endogenously. However, if $Km > \frac{1}{2}$, $\delta(C^*) = 0$. There are no contradictions and gaps at all.

Alternative Restrictions on Interpretations In this subsection, we relax the assumption that I maps from S^C to A, that is, the precedent system and consider the following two alternative restrictions on interpretation.

Firstly, the interpreter's decision is to choose a function $I : [0,1) \mapsto A$ such that: if $a_k = a$ for every $e_k \in E^C$ for which $\omega \in e_k$, then $I(\omega) = a$. That is, the interpreter cannot override an unambiguous clause, but she has total freedom of discretion elsewhere. Under this restriction, the interpreter is not bound by the precedents. Furthermore, there is no need to distinguish between gaps and contradictions, since, in any state in which the interpreter has discretion, the interpreter will choose her own optimal action. Thus what matters to the writer is only the degree of vagueness $\delta(C)$. Of course, contradictions are still possible but not necessary anymore, since they could be replaced by gaps.

Secondly, the interpreter's decision is to choose a function $I : [0,1) \mapsto A$ such that: for all ω, if $J^C(\omega) \neq \emptyset$, then $I(\omega) \in \{a \in A : a = a_k \text{ and } \omega \in e_k \text{ for some } k\}$. This restriction is at the other extreme. When the interpreter faces contradictory clauses, she can only choose one of the actions specified by the clauses. Obviously, there is no rationale for contradictions as well under this restriction.

Thus the contradictory clauses exist necessarily when $Km < \frac{1}{2}$ only under the precedent system.

Comments There are some points worthy to be mentioned.

Firstly, we have compared the results under different restrictions on interpretation method. The optimal restriction on interpretation in the constitutional stage is a natural question to be answered.

Secondly, Sainsbury (1990) defines that a word is *precise* if it describes a well-defined set of objects. By contrast, a word is *vague* if it is not precise. Thus the words "appropriate" and "probably" are vague while the words "leap year" and "male" are precise. Schwartz and Scott (2003) distinguish between ambiguity and vagueness. *Ambiguity* occurs only when at least two distinct, usually inconsistent meanings exist. By these definitions, the contract $C = \{(earthquake, a_1), (fire, a_2)\}$ is only ambiguous, but not vague. It is ambiguous that which particular action should be implemented in the state of both earthquake and fire. But $\{a_1, a_2\}$, the set of actions that could be implemented in the special state, is well-defined.

Therefore, according to these definitions, Heller and Spiegler (2008) endogenize only ambiguity of contracts but not vagueness defined by Sainsbury (1990). Although vagueness is very prevalent in our natural language[40], sometimes real-life contracting parties *purposefully* write vague instructions, say, "taking appropriate care" or "reporting at regular periods". Thus future research could provide a rationale for vagueness in contracts.

Furthermore, the vague instructions, "taking appropriate care" and "reporting at regular periods", are vague actions. Endogenizing not only vague events but also vague actions is also an interesting topic.[41]

Concluding Remarks

The two main models introduced above characterize some aspects of the general problem in different ways. The most important difference is on the method to model the cost of writing contracts. Shavell (2006) is among the literatures that model the writing cost explicitly. Given the writing cost per term, the writer makes an optimal choice of the written contract. However, it is paradoxical that the writer could optimize the contract and thus could take all kinds of terms into account but the writer bears the writing cost only when the final written contract is related to it. It seems that the only writing cost is the physical ink and paper. In contrast to it, Heller and Spiegler (2008) models the writing cost by a limit of linguistic ability of the writer. It is less artificial than Shavell (2006). However, this limit is exogenous in the model. In reality, if the writer bears more writing cost, the complexity of the language written in the contract could be increased. Thus we still need the writing cost to endogenize the degree of complexity of the contract. What we need is a bounded rationality foundation to predict the writer's optimal behavior.

In Chung and Fortnow (2007), there is an interaction between a contract (law) writer and

[40] Beyond the contracting problem, the nature of language vagueness per se is a puzzle. See Lipman (2003).

[41] Battigalli and Maggi (2002) endogenize incompleteness of events and actions, but it belongs to ambiguity not vagueness in this context.

contract interpreter. But they cannot communicate directly. A writer who is unaware and yet aware of her unawareness just writes a simple contract. In equilibrium, simplicity of contracts signals a low awareness of the writer to the interpreter.

Board and Chung (2007) provide a rationale for the legal interpretive doctrine called *verba fortius accipiuntur contra proferentem*, which requires the judge to resolve any ambiguity in a contract against the party who drafted the contract.

3.3 Bayesian Models

Interestingly, Tirole (2009) re-models unawareness and contracts by Bayesian approach.

3.3.1 Costly Cognition and Endogenized Describability

The following setup is considered by Tirole (2009). There are two contracting parties: a buyer (B) and a seller (S). The seller sells a design A to the buyer. The timing is following:

• Stage 1: Nature moves.

Nature selects the *appropriate* design: A with probability $1-\rho$ and A' with probability ρ. The seller's cost of both designs are c. The buyer's payoff of the appropriate design is v and the buyer's payoff of A is $v-\Delta$ if A is not appropriate. But the seller could convert A into A' with zero cost at any stage.

• Stage 2: Cognition stage.

Both the buyer and the seller are uncertain that whether the known A is appropriate or some unknown A' is appropriate.[42] But the buyer could choose b which is the probability that the buyer learns that A' is appropriate under the condition that A' is indeed the appropriate design, but the buyer has to bear a cognitive cost $T(b)$ which is a smooth, increasing and convex function such that $T(0)=0$, $T(1)=\infty$ and $T'(0)=0$. We say a contract is more *complete* if b is higher. That is, the buyer spends more resource on deliberating the design.

• Stage 3: Contracting stage

With probability ρb, the buyer learns that A' is appropriate and communicates it to the seller with no costs. Then they choose the price p_1 such that $v-p_1=(1-\sigma)(v-c)$ where

[42] In other words, A' is not describable ex ante but describable ex post.

$1-\sigma$ is the buyer's bargaining power and, correspondingly, σ is the seller's bargaining power by Nash bargaining solution concept.

With probability $1-\rho b$, the buyer is still uncertain that if A is appropriate or not. So is the seller. They choose the price p_2 such that $v - \hat{\rho}(b^*)h - p_2 = (1-\sigma)(v-c)$.[43] $h < \Delta$ is the value that the seller holds up the buyer in the renegotiation stage when the appropriate A' is revealed.[44] According to Bayesian updating, $\hat{\rho}(b^*) \equiv \frac{\rho(1-b^*)}{1-\rho b^*}$ is their posterior probability that the seller holds up the buyer in equilibrium where b^* is the seller's belief of the buyer's choice of b in equilibrium Thus to avoid being held up by the seller, the buyer spends some transaction cost on cognition.

At stage 2, the buyer solves the following problem by backward induction:

$$\begin{aligned} \max_{b,p_1,p_2} \quad & -T(b) + \rho b(v-p_1) + \rho(1-b)(v-h-p_2) + (1-\rho)(v-p_2) \\ \text{s.t.} \quad & v - p_1 = (1-\sigma)(v-c) \\ & v - \frac{\rho(1-b^*)}{1-\rho b^*}h - p_2 = (1-\sigma)(v-c) \end{aligned} \tag{3.24}$$

Therefore, the equilibrium cognition is b^* such that:

$$T'(b^*) = \frac{\rho(1-\rho)h}{1-\rho b^*} \tag{3.25}$$

From equation (3.25), we see the key insight of Tirole (2007). The equilibrium cognition $b^* > 0$ is increasing in the hold-up problem h. In equilibrium, the buyer's expected payoff is $(1-\sigma)(v-c) - T(b^*)$ and the seller's expected payoff is $\sigma(v-c)$ in stage 1. However, a positive b^* is totally wasteful, since converting from A to A' is costless for the seller. Thus, to avoid being held up by the seller, the buyer spends too much transaction cost to make the contract excessively complete. While standard incomplete contracting literature regards describability problem exogenously, Tirole (2007) endogenizes it.

Moreover, the paper further shows that the over-completeness of the contract is reduced by relational contracting, vertical integration or short term contracting. But ex ante competition does not reduce this cognitive transaction cost.

[43] In Tirole (2007), it is shown that if σ is small enough the buyer is always willing to contract with the seller even when she remains uncertain.

[44] Tirole (2007) assumes that the ex ante and ex post bargaining powers of both the buyer and the seller are the same. Thus $h = \sigma\Delta$. However, we relax this assumption but only restrict h to be less than Δ.

3.3.2 Asymmetric Awareness and Mis-selling

Selling inappropriate products (goods or services) which do not deliver the proper utility levels for consumers - in other words, mis-selling products - is of practical importance in many industries. The problem is especially serious in modern-day industries where consumers are unaware and therefore do not perceive all the payoff-relevant features of the products, unless they incur sufficient cognitive costs before the purchasing decision.

As a result of information disadvantages, consumers may be unaware of some low-quality aspects of products, say, harmful radiation from computer monitors or cell-phones, added chemicals in foods, and side effects of medicines. Consumers may also be adversely surprised by the add-on costs of cartridges after buying a printer, or of using the telephone in a hotel. Moreover, the contractual implications of some services are extremely complex. The Office of Communications in the United Kingdom found mis-selling in telecom services to be a growing problem.[45] Similarly, consumers in financial services are increasingly exposed to the mis-selling of complex financial products, such as endowment mortgages, private pensions, investment funds and insurance products. The Financial Services Authority noted as early as 2000 that one in eight consumers in the United Kingdom who had bought a financial product in the past five years later regretted her choice.[46] More recently, thousands of people in Hong Kong, Singapore and Taiwan took to the streets to protest and demand a refund of the money they lost from the financial products backed by failed Lehman Brothers in the financial crisis.[47]

Economists have noted the importance of the mis-selling problem, but typically emphasize only one particular downside of mis-selling, namely that the consumers are hurt at the post-contractual stage.[48] However, the overall significance of the mis-selling problem is largely underestimated when one only looks at consumer complaints, these being just the "tip of the iceberg". In fact, sometimes mis-selling per se does not reduce social surplus. It hurts the consumer, yet benefits the seller. In these cases, mis-selling is purely a problem of redistribution. In this paper, we shed light on a different and probably more important problem of mis-selling: the transaction cost of pre-contractual cognition by consumers, which is the "submerged part of the iceberg".

Hayek (1945) argues that not only does the price mechanism help to utilize knowledge

[45] See *Protecting citizen-consumers from mis-selling of fixed-line telecoms services,* Office of Communications, UK, 22 November 2004.

[46] See *Informed decisions? How consumers use Key Features: a synthesis of research on the use of product information at the point of sale,* Financial Services Authority, November 2000.

[47] See the article "Troubled Securities in Asia" in *The Economist*, November 20th 2008.

[48] For example, Inderst and Ottaviani (2008) study the mis-selling problem in the principal-agent framework. Furthermore, their paper focuses on the intra-organizational incentives for an employee not to mis-sell, while our paper sheds light on the direct interaction between a buyer and a seller in the market.

dispersed among individuals, but it also promotes the efficient division of knowledge. When consumers have unforeseen contingencies, however, the price mechanism may fail. If the regulator also lacks the corresponding knowledge and therefore cannot promote the awareness of consumers, mis-selling prevents the division and sharing of knowledge among individuals. Although consumers may be unaware, they are aware that they may be unaware of something. The possibility of mis-selling induces too much cognitive resources being spent on pre-contractual thinking by the consumers in order to avoid their being potentially hurt. To return to the examples mentioned above, in order to avoid being exploited, consumers have to read many books, surf the Internet, or discuss with people specialized in computer technology, food safety, telecommunications, or finance, which results in socially wasteful duplication of cognitive efforts.

In this paper, we address this issue through a model along the lines of Tirole (2009). We study the interaction between a seller (he) and a buyer (she) who is aware of the possibility that the status quo product may not be appropriate, meaning that at the post-contractual stage she may not be satisfied with it. However, the buyer can think about the appropriateness of the product before contracting.

We deviate from Tirole (2009) in two main aspects.[49]

First, we assume that the seller knows whether the status quo product is appropriate for the buyer or some novel product, which is unforeseen by the buyer, is appropriate, whereas in Tirole's model both parties are uninformed *ex ante*. The assumption that the seller is better informed is natural in the mis-selling context because of the seller's specialization in the industry. Given his knowledge, the seller can strategically announce the appropriateness of the product to the buyer. If the status quo product is not appropriate, the seller either mis-sells it or offers an alternative (novel) product that opens the buyer's eyes. After introducing asymmetric awareness between the seller and the buyer, in this paper, we focus on the problem of strategic mis-selling.

Second, in Tirole's model, the seller is uninformed *ex ante*, so selling inappropriate products is not the seller's deliberate action, and there is no need to examine instruments that deter him from doing so. In contrast, in our model, we allow a transfer of money from the seller to the buyer in the case of mis-selling. This can be the result of a litigation process. Since it is commonly known that the seller knows the product-appropriateness, the transfer is to deter the seller's intent of mis-selling.

In our model, three key parameters determine the equilibrium. The first parameter is

[49]Besides these two differences, we assume that the buyer has the full bargaining power for analytic simplicity. Moreover, we abstract away from the adjustment cost in Tirole's model of renegotiation by assuming that mis-selling does not directly reduce social surplus. The role of a positive adjustment cost in his model is to produce the interesting result of the buyer's insufficient cognition as a free riding problem. However, since the seller is informed about the product-appropriateness in our mis-selling problems, the pre-contractual cognition of the buyer is purely rent-seeking and therefore wasteful.

the *a priori* probability that the status quo product is not appropriate, which may lead to mis-selling. We call this the *extent* of the mis-selling problem. The second parameter is the loss of the buyer when the seller successfully mis-sells. We call this the *effect* of the mis-selling problem. The last parameter is the transfer of money from the seller whose mis-selling behavior is found out by the buyer.

We show that there is no *separating* equilibrium in which the seller always truthfully reports the appropriate product. In the case where the status quo product is inappropriate, if the seller does not mis-sell, that is, the seller reveals the novel product, the buyer has no incentive to question the purchasing decision. However, it jeopardizes the seller's incentive to report truthfully, as mis-selling can never be found out before contracting.

When the extent of mis-selling is low in the economy as in many developed countries, there is a *pooling* equilibrium where the seller always announces that the status quo product is appropriate, even though it may not be appropriate. The buyer has a low incentive to question the purchasing decision, thus the seller prefers to mis-sell as the probability of being punished without getting any rent is low. Therefore, developed countries produce "simple men".

When the extent of mis-selling is high as in some less-developed countries, we have no pure-strategy equilibrium. If the seller mis-sells with certainty whenever it is possible, the buyer has to stay on her toes so as to avoid the mis-selling. The seller then has no incentive to mis-sell, since the probability of being caught is high. Conversely, the argument for no separating equilibrium applies. Thus, given that the status quo product is inappropriate, only when the seller randomizes between truth-telling and mis-selling appropriately, can the buyer choose a corresponding cognition level such that seller is indeed indifferent between truth-telling and mis-selling, which generates a *semi-separating* equilibrium. This is in contrast to Tirole (2009) where the seller would strictly prefer to mis-sell. Further, it is worth noting that, in these countries, as the equilibrium cognition level is high, consumers are relatively know-it-all.

We define the transaction cost as the expected cognition cost of the buyer. We show that in the pooling equilibrium, the transaction cost is increasing in the extent of the mis-selling problem. In a semi-separating equilibrium, however, we have the *opposite* result: the greater the extent of the mis-selling, the smaller the transaction cost, because a greater extent of mis-selling induces a much higher probability of awareness-inducing information disclosure by the seller. Hence, the transaction cost is increasing in the extent of the mis-selling problem as long as the extent of the mis-selling is small as in many developed countries, and is decreasing thereafter, as in some less-developed countries.

Note that only in a separating equilibrium does the transaction cost vanish, because, in this case, there is no mis-selling and the buyer therefore does not think. However, the separating equilibrium does not exist. Thus, our ideal society where worthy trust is everywhere and there is no wasteful cognition is an impossible world in the one-shot interaction setting. As "the economic institutions of capitalism have the main purpose and effect of economizing on transaction costs", (Williamson, 1985, p. 17) we then study whether the particular transaction cost of pre-contractual cognition can be saved by some properly designed market institutions.

First, we show that reputation may induce a separating equilibrium in each period, which shatters the mis-selling behavior of the seller, if both buyer and seller are sufficiently patient and the probability of detecting past cheating behavior is sufficiently high. In the separating equilibrium, the buyer rewards the seller's truthful report of the novel product by paying a higher price, or a "tip". However, whenever mis-selling occurs in some period in the separating equilibrium, the economy cannot go back to the outcome where no pre-contractual cognition of the buyer is exerted no matter how long the policy of anti-mis-selling campaigns has been exercised. But the out-of-equilibrium amnesty, which destroys the evidence of this mis-selling spot, can let the economy return immediately to the outcome where the transaction cost vanishes. Thus, the "big-bang" approach of reform is more effective than the gradualist strategy in the context of mis-selling problems. Lastly, if it is impossible for the seller to know the product-appropriateness in some period, we cannot save the pre-contractual cognition completely. Nonetheless, a smaller probability of unawareness of the seller reduces the transaction cost, which reflects the value of the seller's specialization.

Second, in the one-shot interaction setting, if two sellers compete for the buyer simultaneously, there is always a pure-strategy equilibrium where both sellers truthfully report the appropriate product. In contrast to Tirole (2009), in the context of asymmetric information, competition between sellers plays a role of reducing transaction costs, as two sellers may simultaneously reveal the novel product to the buyer. Hence, reputation or competition may induce a separating equilibrium and minimize the transaction cost.

The remainder of the paper is organized as follows: Section 3.3.2 reviews the related literature. Section 4.3.2 presents the basic model. Sections 3.3.2 and 3.3.2 discuss the role of reputation and competition, respectively. We conclude in section 4.3.5.

Related literature

Economics of Information This paper considers a novel contracting problem with asymmetric information.[50] First, at the contracting stage, the uninformed party, namely the buyer, proposes the price (in each bargaining period, and uses delay as a screening device as in Appendix A.1.12). This problem belongs to the standard screening problem,

[50]Cf. Bolton and Dewatripont (2005) and Laffont and Martimort (2002) for a detailed account of this extensive literature on economics of asymmetric information.

which was first formally developed by Mirrlees (1971). Second, our model is close to the price-quality-signaling literature especially by Bester and Ritzberger (2001), which endogenizes the buyer's information on the product quality by the information acquisition approach introduced by Hirshleifer (1971), and Crémer, Khalil and Rochet (1998a, b). The pre-contractual cognition in our model is a special kind of information acquisition in the unawareness context. However, the buyer has the full bargaining power in our model. Thus, we abstract from the signaling problem, yet focus on the cognition-saving mechanisms. Furthermore, in our model, the seller *may* be able to announce some awareness-inducing information that opens the buyer's eyes, which is absent in the standard literature of asymmetric information.

Incomplete Contracts We have several rationales for incomplete contracts so far: non-verifiability (Williamson (1985), Grossman and Hart (1986), Hart and Moore (1990), Aghion and Tirole (1997), Maskin and Tirole (1999)), signaling (Aghion and Bolton (1987), Spier (1992), Hermalin and Katz (1993)), explicit writing costs (Dye (1985), Anderlini and Felli (1999), Battigalli and Maggi (2002)), reading costs (Rasmusen (2001)), and second-best incentives: e.g., effort substitution in multi-task environments (Holmström and Milgrom (1991)) and strategic ambiguity (Bernheim and Whinston (1998)).

Recent approaches endogenize incompleteness of contracts from bounded rationality introduced by Simon (1957). Bolton and Faure-Grimaud (2009) and Tirole (2009) endogenize incomplete contracts from the parties' insufficient cognition. In contrast, when one party is fully rational and the other party is boundedly rational, contractual incompleteness can be the result of strategic shrouding by the rational party. (Filiz-Ozbay (2007), Gabaix and Laibson (2006), von Thadden and Zhao (2009)) In our paper, the incomplete contract, which leads to mis-selling, stems from both the buyer's inadequate cognition and the seller's strategic shrouding.

Unawareness and Contract Design In cognitive psychology, Kahneman and Tversky (1973, 1974) argue that people use an availability heuristic to judge probabilities. In economics, the corresponding notion is unawareness.

Fagin and Halpern (1988) and Modica and Rustichini (1994) first discuss unawareness formally. Later, Heifetz, Meier and Schipper (2006), Li (2009), and Galanis (2007) model unawareness, which circumvents the impossibility result of non-trivial unawareness by Dekel, Lipman and Rustichini (1998). From then on, unawareness is possible to express. Its economic implications in contracting problems are discussed in the following papers. Gabaix and Laibson (2006) model consumers' unawareness of some add-ons (actions of one's opponents). Eliaz and Spiegler (2006) study screening consumers' awareness of future changing tastes (preferences). von Thadden and Zhao (2009) design the incentives for agents who are possibly unaware of their own choice possibilities (one's own actions). Filiz-Ozbay (2007) examines an insuree's unawareness of future contingencies of calamities (actions of nature).

Considerable progress in modeling unawareness has been made by Board and Chung (2006), which makes awareness of unawareness possible of expression.[51] Awareness of unawareness plays an important role particularly in contracting problems. Although contracting parities are unaware, they are aware that they might be unaware of something. For example, Chung and Fortnow (2008) model the interaction between a contract writer and an interpreter. In equilibrium, the writer who is aware that she is unaware of something writes a simple contract to signal her low awareness. In addition, awareness of unawareness also plays an important role in Tirole (2009) and the present paper. A contracting party exerts cognitive efforts before contracting only if he is aware of his potential unawareness.

The Basic Model

The Setup There are two risk-neutral contracting parties: a buyer (B, she) and a seller (S, he). The buyer wants to buy one unit indivisible product (good or service) from the seller. There is a *status quo* product A available from the seller. A can be interpreted as a displayed good on sale or a standard form contract for a service. If A is appropriate, A delivers utility $v > 0$ to the buyer. If A is not appropriate, however, the buyer's utility from A reduces to $v - h$ with $h > 0$. In the latter case, we assume that there always exists a *novel* product A', which is appropriate, (meaning it delivers utility v to the buyer,) but is unforeseen by the buyer.

If A is not appropriate, mis-selling may come in many guises. First, in a *low-cost-low-quality* interpretation, the buyer may be unaware of some utility-relevant features of the status quo product before contracting. For example, after using a computer monitor with a severe radiation problem (A), the buyer is hurt by h, but this monitor is h-cheaper to produce than a LCD monitor (A'). Second, in an *add-ons* interpretation, as in the example in Gabaix and Laibson (2006), the buyer may be unaware of the future cost of cartridge when buying a printer. If the seller sells a status quo printer with high future cartridge costs, and shrouds the attribute of cartridge to the buyer, the seller saves a cost h for the ink-saving technique. But the buyer has to pay h for the cartridge in the future. By these two interpretations, we normalize only the cost of producing the appropriate product to zero, whereas the seller's cost of producing A is $-h$ if A is not appropriate. Finally, in a *hold-up* interpretation in the style of Tirole (2009), the unforeseen features of the product, say a design, are not costly for the seller to adjust

[51] In dynamic games, Halpern and Rego (2006) and Heifetz, Meier and Schipper (2009) also discuss awareness of unawareness. Further, Chen and Zhao (2009) synthesize the existing solution concepts of contracting models with awareness of unawareness in a principal-agent framework.

ex post. That is, if A' is appropriate, the seller could convert A into A' at zero cost. At the post-contractual stage, however, the buyer is locked in, so the seller can hold her up by asking for more payment h. In other words, the seller blackmails the unaware buyer *ex post*. By this interpretation, the seller's cost of producing any product is normalized to zero.

Hence, the *effect* of the mis-selling problem is modeled by a constant h. If A is not appropriate, and the buyer consumes A, the magnitude h is not only the buyer's loss from mis-selling but also the rent for the seller who mis-sells. By mis-selling, there is a redistribution of payoff from the buyer to the seller.[52]

Figure 3.6 shows the time line, while Appendix A.1.11 presents the detailed game structure.

The more detailed timing is as follows:

• Stage 1: Nature moves.

Nature N chooses A to be *not* appropriate with probability ρ. In other words, with probability ρ, something of the status quo product hurts the buyer. We call ρ the *extent* of the mis-selling problem, since, with probability ρ, the status quo product is inappropriate, which leads to the potential mis-selling problem.

In contrast to Tirole (2009), we assume that the seller knows nature's move, while the buyer does not.[53] Facing the known product A, the buyer is aware that A might be not appropriate for her. In other words, the buyer is aware that something may go wrong with A. Here, we assume common knowledge of the game and rationality. Since the two parties have a common prior ρ, the problem is a classical one of asymmetric information between the buyer and the seller. If A is appropriate, we call the seller type-A; otherwise, we call him type-A'.

• Stage 2: Seller's announcing stage.

[52] In general, the gain of the seller in the case of mis-selling h' may be different from the loss of the buyer h. Particularly, it is quite plausible to assume a deadweight loss of mis-selling ($h' < h$). This extension is straightforward without changing the qualitative results. Our motivation of this simplification compared with Tirole (2009) is that since the seller has been already informed about the product-appropriateness, the buyer's pre-contractual cognition is purely rent-seeking. Therefore we focus only on this more important transaction cost in the mis-selling context, and rule out the possibility of insufficient cognition of the buyer as a free riding problem in Tirole (2009) which needs the assumption of a deadweight loss of hold-up.

[53] The seller's information advantage is common in situations where the seller is a specialized firm that has the capacity to design the product and thus gains detailed knowledge of the product-appropriateness, whereas the buyer who demands something and has no ability to produce it is therefore relatively unfamiliar with product. Nevertheless, we get back to the case where the seller may be uninformed as well in section 3.3.2.

At this stage, the seller can say something to the buyer. He may announce that the status quo product A is appropriate, or point out that A does not deliver v to the buyer and shows a novel product A', which is appropriate.

If the seller is type-A, then he can only announce that A is appropriate, since A is indeed appropriate, and the buyer cannot be harmfully surprised *ex post*. However, type-A' seller contemplates two options: falsely saying that the status quo product A is appropriate (mis-selling) on the one hand and unveiling A' (truth-telling) on the other hand.

Note that we have an asymmetry of information transmission in two states of nature (A and A'). Here, announcing A provides only *soft* information. The buyer remains uncertain of the appropriateness of A (if the seller may mis-sell in equilibrium). However, announcing A' immediately reveals the *hard* evidence that A is not appropriate, and is therefore an eye-opener for the buyer, because the seller can report A' to the buyer only if A is de facto not appropriate. Intuitively, the seller can announce something surprising only if the buyer is indeed unaware of something. For example, if the buyer is unaware of the future cartridge when buying a printer, the seller may hide the cost of cartridge or mention it to the buyer. In other words, if nothing of the status quo product goes wrong, the seller cannot prove it just by words (which are cheap). However, if something of the status quo product is really wrong, the seller can provide the awareness-inducing information to the buyer.[54]

Let q be the probability that type-A' seller mis-sells, which is endogenous.

• Stage 3: Buyer's cognition stage.

If $q > 0$ and the seller says A is appropriate, the buyer still does not know whether or not A is appropriate. However, the buyer can think, that is, make an effort to contemplate the situation in this case. For example, an investor can read a book of finance for "dummies", surf the Internet to search for the relevant financial information, or discuss with friends who are familiar with finance.

Formally, the buyer chooses her *cognition* level b, which maps bijectively to the probability that the buyer learns that A' is appropriate given that A' is in fact the appropriate product. In other words, if something of the status quo product goes wrong, the buyer can find it out with probability b. However, if nothing goes wrong (A is appropriate) the buyer finds nothing after pre-contractual cognition.

The buyer bears a cognition cost, or thinking cost, $C(b)$ that is a smooth, strictly

[54] The feature of asymmetry of information transmission in two states (A and A') here is thus different from cheap-talk games (See, e.g., Crawford and Sobel, 1982) where information is always soft and persuasion games (See, e.g., Milgrom (1981) and Milgrom and Roberts (1986)) where information is always hard.

increasing and strictly convex function with the properties $C(0) = C'(0) = 0$ and $C(1) = \infty$.

• Stage 4: Contracting stage.

We assume for simplicity that the buyer has full bargaining power at the contracting stage and thus makes a take-it-or-leave-it offer to the seller. Nevertheless, Appendix A.1.12 shows that the pricing result is robust to a more general bargaining protocol where the buyer makes all the offers in an infinite-horizon setting. We also assume the outside options of both parties yield zero payoffs.

If the seller reveals A' to the buyer, the buyer is suddenly aware of A' and proposes a contract that consists of a price p and a full specification of the product A'.[55] Facing the contract, the seller decides whether to accept it.

If the buyer is told that A is appropriate, and after the buyer has thought about it at stage 3, there are two possibilities.

(1) The buyer learns that A is not appropriate by pre-contractual cognition, thus she knows the seller has cheated her. The buyer then demands A' and proposes a price p subject to the seller's participation.[56] We assume that whenever the buyer knows that the seller mis-reports the product, the buyer can sue the seller for a monetary transfer $t \in (0, h)$.

To shed light on the mere transaction cost of pre-contractual cognition, we assume that the buyer receives the transfer even without consuming the product. In fact, what is crucial in the model is punishing the cheating type-A' seller so as to deter mis-selling even if the buyer has not consumed the product yet. To avoid the welfare loss from punishing the seller, the monetary punishment should be returned to society, namely, the buyer, as she is the only remaining player in the game. We consider the case of multiple buyers in Appendix A.1.13 where an individual buyer's cognition may be reduced as a free-riding result.

Since we have assumed that there is common knowledge that the seller knows the product-appropriateness, the buyer knows that mis-selling happens not by chance but occurs only if the seller has an intent of doing it. Hence, the buyer will sue the seller

[55] Of course, after the seller reveals A', the cost of understanding A' for the buyer is not zero in reality. However, it should be much lower than the cost of learning A' by the buyer alone. Thus, without loss of generality, we normalize the cost of understanding A' to zero. Further, the seller may manipulate the understanding cost for the buyer, say, by disclosing the awareness-inducing information only in the fine print. In this case, we would rather interpret the seller's behavior as strategic mis-selling (announcing A).

[56] Introducing the possibility that the buyer may demand A does not change our results. In equilibrium, the buyer will gain the same share of the surplus in trading both A' and A, but the equilibrium price p is lower in the latter case.

for willful mis-selling. Even if it is possible that the seller does not know the product-appropriateness, due to the seller's specialization in the industry, we assume that the seller can acquire the information with zero cost. Hence, liability of the seller requires the seller to take appropriate care to report the product.[57] As Milgrom (2008) argues

> "... what is needed ... is to hold the seller liable for failures to reveal promptly not only the verifiable information that the seller knew, but also the information that it *should have known* under the circumstances."

Now one might consider increasing the transfer t arbitrarily in the constitution so as to deter mis-selling. In reality, however, the judicial system is imperfect. When the seller mis-sells, the court will judge it correctly with probability z. Because of the seller's limited liability W, it is very likely that the *expected* transfer t is less than zW.[58]

It is worth noting that a deterministic transfer t in the case of mis-selling can be implemented via a litigation mechanism as shown in Appendix A.1.14 when only type-A' seller can provide hard evidence of mis-selling to the court of law. When there is symmetric information of mis-selling between the buyer and the seller *ex post*, the transfer can be realized based on the *spirit* but not the *letter* of their contract, although product-appropriateness is not verifiable by the court of law. We also show in Appendix A.1.14 that even in this case t cannot be arbitrarily high. Here we content ourselves with a simple analysis in which the transfer is modeled by an exogenous constant t smaller than h whenever the buyer knows the seller's false report.

(2) The second possibility is that the buyer remains uncertain of whether A is appropriate. Then the buyer proposes a contract, including a price p and the specification of A, under uncertainty.

Suppose A' is appropriate. The buyer can only demand A, although it is not appropriate. The imperfect description in the contract reflects a misleading contract. If the seller

[57] A similar discussion is in the accident (tort) law literature. Here the buyer is a victim, and the seller is an injurer. Assume that the mis-selling accident is unilateral, namely only the seller's care affects the mis-selling risk. We impose the rule of strict liability, meaning the seller has the liability to pay all the loss for the buyer within the seller's asset requirement. Note that the negligence rule, which says that the seller pays for the buyer's loss only if the seller's care is less than the level the court specifies (due care), is not feasible when the court cannot detect the seller's efforts of care. (See, e.g., Shavell, 2004)

[58] One may wonder whether we can solve the problem by the standard Nash implementation approach by Maskin (1999) or subgame perfect implementation approach by Moore and Repullo (1988) and Maskin and Tirole (1999). However, both approaches cannot provide unique equilibrium, as the preferences of both parties over the transfer decision are state-independent, namely, the buyer always prefers a transfer, and the seller always prefers no transfer.

One may also wonder whether we can solve the problem by these approaches before the buyer's uncertainty is resolved. However, standard approaches are not robust to the asymmetric information context as we assume here. See, e.g., Aghion, Fudenberg and Holden (2009) and Kunimoto (2008).

accepts the contract, then, at the post-contractual stage, the buyer receives payoff

$$v - p - h + t - C(b),$$

and the seller receives

$$p + h - t.$$

Ex post, the buyer will find out that A is actually not appropriate, because she is hurt by h. The buyer then sues the seller for a transfer t.

Suppose A is appropriate. Since there is no mis-selling problem, if the seller accepts the contract, then, at the post-contractual stage, the buyer receives $v - p - C(b)$, and the seller receives p.

Equilibrium We solve the game backwards by using the solution concept of perfect Bayesian equilibrium.[59]

Pricing At stage 4, when type-A' seller tells the buyer A' or the buyer finds that A' is appropriate by cognition, the buyer optimally demands A'. Sequential rationality implies that the seller accepts the contract if and only if $p \geq 0$ and the optimal price proposed by the buyer is $p = 0$.

An interesting finding is that when the seller says A is appropriate and the buyer finds nothing after cognition the optimal price remains $p = 0$. To show this point, let

$$\widehat{\rho} \equiv \frac{\rho q(1-b)}{1 - \rho + \rho q(1-b)}$$

be the posterior probability of mis-selling from the buyer's view given that the buyer finds nothing after cognition and the buyer believes that type-A' seller's probability of mis-selling is q.[60] Here the buyer's belief is updated according to Bayesian rule.

Suppose the buyer proposes some price $p \geq 0$. Then both types of sellers accept it. The buyer therefore receives her expected payoff

$$U_1 \equiv (1 - \widehat{\rho})(v - p) + \widehat{\rho}(v - h + t - p) - C(b).$$

The buyer's best proposal is $p = 0$ given that $p \geq 0$.

[59] The problem is modeled as a standard extensive form game. The connection to games with unawareness is discussed in Appendix A.1.15. In general, games with unawareness can be mapped to games with incomplete information, and the extended solution concepts of the former coincide with the corresponding standard solution concepts of the latter. See, e.g., Feinberg (2009).

[60] In equilibrium, the buyer has a correct belief of q.

Suppose the buyer proposes price $p \in [t-h, 0)$. Then type-A seller will reject it. The buyer receives her expected payoff

$$U_2 \equiv \widehat{\rho}(v - h + t - p) - C(b).$$

The buyer's best proposal is $p = t - h$ given that $p \geq t - h$.

If $p < t - h$, both types of sellers reject it. Thus, the buyer receives payoff $-C(b)$.

To focus on the welfare loss of pre-contractual cognition, we abstract from the inefficient contracting result where a mutually beneficial trade may break down. Now we make the following assumption:

Assumption 3.5 $v - \rho\,(h - t) > \rho v$.

The interpretation of Assumption 3.5 is, given that there is no cognition and type-A' seller mis-sells with certainty, the buyer always prefers $p = 0$ and contracting with both types of sellers (where the buyer gets $v - \rho\,(h - t)$) to $p = t - h$ and contracting with only type-A' seller (where the buyer gets ρv). Hence, we rule out the case in which there are only "lemons" in the market, and mutually beneficial trades break down for type-A seller. (Akerlof, 1970) This assumption holds when the extent and the effect of mis-selling are not too high, namely ρ and h and are sufficiently small, and the gain from trade v is large enough.

Assumption 3.5 implies

$$(1-\rho)v - \rho\,(h-t) > 0$$

that also implies

$$(1-\widehat{\rho})v - \widehat{\rho}\,(h-t) > 0, \tag{3.26}$$

since $\widehat{\rho} < \rho$ holds for all $b > 0$ and q.

Inequation (3.26) is equivalent to $U_1 > U_2$. Hence, $p = 0$ is better than $p = t - h$ for the buyer. Intuitively, after the buyer exerts cognitive efforts and finds nothing, and type-A' seller *may* not mis-sell with probability one, the buyer believes that mis-selling is less likely to occur, and is more willing to contract with both types of sellers.

Moreover, Assumption 3.5 also implies that the price $p < t - h$ is never optimal, as $v > 0$. For the buyer, contracting with both types of sellers is also better than her outside option, since there is a positive gain from trade.

Furthermore, Appendix A.1.12 shows that if Assumption 3.5 holds, $p = 0$ without delay is also the equilibrium outcome in a general bargaining game where the buyer makes all the offers in an infinite-horizon setting in which delay can work as a screening device for the buyer.[61]

[61] In Appendix A.1.16, we show that if this assumption fails, there is even no equilibrium in which the buyer strictly prefers to contracting with only type-A' seller or proposing nothing.

We now continue to solve the game backwards.

No Separating Equilibrium Suppose, at stage 2, $q = 0$, i.e., type-A' seller tells the buyer A' with certainty. Then at stage 3, the buyer's optimal cognition is $b = 0$. Since there is no mis-selling problem anymore, it is not worthwhile for the buyer to spend any resource on thinking. However, if $b = 0$, it turns out that type-A' seller optimally pretends to be type-A, since type-A' seller will get the rent from mis-selling with certainty (and $h - t > 0$). Thus, it is impossible to have a *separating equilibrium* in which the buyer can tell type-A seller and type-A' seller apart for sure without cognition.[62]

Pooling Equilibrium Alternatively, suppose $q = 1$ at stage 2, i.e., type-A' seller mis-sells with certainty. Then at stage 3, given that $p = 0$, the buyer maximizes her payoff in expectation

$$\max_b (1-\rho)v + \rho b\,(v+t) + \rho(1-b)(v-h+t) - C(b)$$

where $1 - \rho$ is the probability that A is appropriate, ρb is the probability that A is not appropriate and the buyer knows it by pre-contractual cognition, and $\rho(1 - b)$ is the probability of mis-selling.

The assumptions on $C(\cdot)$ imply that the optimal cognition is b^* such that

$$C'(b^*) = \rho h. \tag{3.27}$$

The marginal cost of cognition equals the marginal benefit from avoiding mis-selling. Equation (3.27) reflects that the equilibrium cognition is increasing in ρh, namely, the product of the extent and the effect of the mis-selling problem.

When ρ is small, we have that b^* is small. Thus, if t is also not too large, we have that

$$(1-b^*)(h-t) + b^*(-t) \geq 0, \tag{3.28}$$

which means that $q = 1$ is indeed optimal for type-A' seller. Therefore, when ρ and t are small, there is a *pooling equilibrium* in the sense that both type-A and type-A' sellers announce that A is appropriate. Intuitively, if the extent and the transfer from type-A' seller are low, the buyer therefore does not exert too much cognition, then type-A' seller has an opportunity to mis-sell.

[62] The negative result is akin to the Grossman-Stiglitz paradox, which says that there is no pure-strategy equilibrium of pricing when acquiring quality-information is costly for consumers in the market. (See Grossmann and Stiglitz, 1980)

Formally, if $\rho h \leq C'(1-\frac{t}{h})$ holds, we have such pooling equilibrium.[63] It is straightforward to see that pooling is more likely to occur for smaller ρ and t, yet the role of h is indeterminate. It is so because increasing h raises the benefit of mis-selling for the seller, which enhances the seller's incentive to mis-sell, yet also raises the buyer's cognition level, which reduces the seller's incentive to mis-sell. Hence, we cannot judge its impact on the validity of condition (3.28).

Semi-Separating Equilibrium However, when $\rho h > C'(1-\frac{t}{h})$, there is no pure-strategy equilibrium, since in this case we have

$$(1-b^*)(h-t)+b^*(-t)<0, \tag{3.29}$$

i.e., a high level of cognition by the buyer in the pooling equilibrium deters type-A' seller from mis-selling.

For large ρ and t, let us therefore investigate the mixed-strategy equilibrium. We assume all parties have perfect recall, so the concept of behavioral strategies is tantamount to that of mixed strategies. Suppose type-A' seller chooses a behavioral strategy $q \in (0,1)$ at stage 2. Then at stage 3, the buyer solves the following problem upon observing A:

$$\max_b \frac{(1-\rho)\,v+\rho qb\,(v+t)+\rho q(1-b)(v-h+t)}{1-\rho+\rho q} - C(b)$$

where the posterior probability that A is appropriate is $\frac{1-\rho}{1-\rho+\rho q}$, the posterior probability that A' is appropriate and the buyer knows it by cognition is $\frac{\rho qb}{1-\rho+\rho q}$, and the posterior probability of mis-selling is $\frac{\rho q(1-b)}{1-\rho+\rho q}$.

Similarly, our assumptions on $C(\cdot)$ guarantee that the optimal cognition is b^* such that

$$C'(b^*)=\frac{\rho qh}{1-\rho+\rho q}. \tag{3.30}$$

[63] Pooling equilibrium occurs if and only if

$$(1-b^*)(h-t)+b^*(-t)\geq 0$$

where b^* is characterized by equation (3.27), which is equivalent to

$$b^* \leq 1-\frac{t}{h}.$$

By strict convexity of $C(\cdot)$, it is also equivalent to

$$C'(b^*) \leq C'(1-\frac{t}{h}),$$

which is nothing but

$$\rho h \leq C'(1-\frac{t}{h}).$$

Since type-A' seller plays a non-degenerate behavioral strategy, at stage 2, he is indifferent between announcing A and A', i.e., the following equation is satisfied.

$$(1 - b^*)(h - t) + b^*(-t) = 0.$$

Therefore, the equilibrium cognition is

$$b^* = 1 - \frac{t}{h}, \tag{3.31}$$

so the equilibrium cognition level is determined by h and t and independent of ρ. In particular, b^* is increasing in h and decreasing in t. If h is higher, type-A' seller has a higher incentive to mis-sell. To keep the seller still indifferent between announcing A and A', the buyer has to think more carefully to reduce type-A' seller's incentive to mis-sell. We have the opposite and yet analogous intuition for a higher t.

Plugging b^* in (3.31) into equation (3.30), we have

$$q^* = \frac{(1-\rho)C'(1-\frac{t}{h})}{\rho(h - C'(1-\frac{t}{h}))}. \tag{3.32}$$

In equilibrium, an appropriate q^* induces the buyer to choose the optimal cognition such that type-A' seller is indeed indifferent between mis-selling and truth-telling, which we dub *semi-separating equilibrium.*[64] This result is in contrast to Tirole (2009). Although the seller has no bargaining power, in the presence of the transfer t, type-A' seller here has no strict incentive to shroud A'. Disclosure of A' by the seller is possible, though not necessary, in equilibrium.

We summarize the results we have so far in the following proposition.

Proposition 3.8 *Under Assumption 3.5, the equilibrium price is* $p = 0$ *that is accepted by the seller.*

There is no separating equilibrium where two types of sellers are distinguished by the buyer with certainty.

[64] The feature of semi-separating equilibrium smacks of an inspection game. The buyer is the counterpart of an inspector, and the seller is the counterpart of an inspectee. (See, e.g., Avenhaus, von Stengel and Zamir, 2002) Nevertheless, there are some substantial differences. First, in our model, there are heterogeneous sellers, and it is impossible for type-A seller to mis-sell. Hence, when the fraction of type-A seller, namely ρ, is high, there exists a pure-strategy equilibrium that is a pooling one. Second, in most inspection games, the inspector has a binary choice: alarm or not alarm. But our buyer here chooses a continuous cognition level b to update her belief. Thus, the buyer chooses an appropriate b, which is a pure strategy, in equilibrium. Third, in inspection games, players move simultaneously, yet in our model, the seller moves first. If the seller reveals A', the buyer therefore needs not to think. Fourth, there are two types of errors for the inspector in the statistical parlance. However, our model excludes the buyer's type I error, since describing A' implies that A' is indeed appropriate.

If $\rho h \leq C'(1 - \frac{t}{h})$, there is a pooling equilibrium where both types of sellers announce that A is appropriate, and the buyer's cognition b^* is characterized by equation (3.27).

Otherwise, there is a semi-separating equilibrium where type-A' seller randomizes between mis-selling and truth-telling with type-A' seller's probability of mis-selling q^* given by equation (3.32), and the buyer's cognition level b^* is given by equation (3.31).

Graphical Interpretation Figures 3.7 and 3.8 illustrate the buyer's best response function $b(q)$ and type-A' seller's best response correspondence $q(b)$ in semi-separating equilibrium and pooling equilibrium, repetitively. Here $b(q)$ is strictly increasing in q,[65] $b(0) = 0$, and $b(1) < 1$ by equation (3.30). Inequations (3.28) and (3.29) imply $q(b) = \{1\}$ for small b, $q(b) = \{0\}$ for large b, and $q(b) = [0, 1]$ for some b in-between.

Figure 3.7 shows that if q is too large, the buyer chooses a large $b(q)$ to avoid being mis-sold to. But a large b implies that $q(b)$ is zero, i.e., the seller dare not to mis-sell. Conversely, if q is too small, the buyer has no incentive to think and thus chooses a small b. A small b implies $q(b)$ is one, i.e., the seller has incentive to mis-sell. Thus, the equilibrium q^* has to be not too large and not too small. In the figure, the intersection of $b(q)$ and $q(b)$ is the unique candidate for equilibrium b^* and q^*.

Thus, when ρ and t are large, there is a semi-separating equilibrium in the sense that type-A' seller randomizes between mis-selling and truth-telling. In contrast, when ρ and t are small, the pooling equilibrium is depicted in Figure 3.8. In this case, $b(q)$ and $q(b)$ intersect at $q = 1$, which implies a pure-strategy equilibrium.

However, since $b(q)$ is strictly increasing in q and b is positive in any equilibrium, $b(q)$ and $q(b)$ can never intersect at $q = 0$. Thus, separating equilibrium does not exist, as the figures show.

In summary, Figure 3.9 illustrates the results in Proposition 3.8.

Economic Implications Pooling equilibrium happens in many developed countries where unforeseen contingencies are not so likely to occur. In these countries, the consumers are "simple" men in the sense that they do not acquire much information on the product-appropriateness. Although the seller mis-sells with certainty whenever he has an opportunity, the consumers are not afraid of it, because it is a small-probability event.

[65] By equation (3.30), we infer that

$$\frac{dC'(b^*)}{dq} = \frac{d\left(\frac{\rho q h}{1-\rho+\rho q}\right)}{dq} = \frac{\rho(1-\rho)h}{(1-\rho+\rho q)^2} > 0.$$

However, semi-separating equilibrium happens in less-developed countries. As the extent of mis-selling is high, the consumers have to be alert. Thus, the consumers in less-developed countries are relatively less simple. Nevertheless, in semi-separating equilibrium, the seller may open the consumers' eyes. The consumers therefore need not acquire too much information like a "know-it-all".

Robustness The results so far are robust to a number of extensions.

Heterogeneous Buyers: We assume only partially unaware buyers, meaning the buyers are aware of their potential unawareness. In reality, a number of buyers are completely unaware, i.e., they are naive in the sense that they always believe what the seller says, and do not exert any cognition. If we allow diversely unaware buyers, the results are qualitatively similar. The main change is that the pooling equilibrium is more likely to occur, because type-A' seller has a higher incentive to mis-sell due to the opportunity to exploit more completely unaware buyers in the population.[66]

Furthermore, if we allow that a fraction of buyers have the same information as the seller has, that is, some buyers know A' without the need of costly cognition when A is not appropriate, the qualitative results remain as well, except that the pooling equilibrium is less likely to occur. More informed buyers in the population increase the probability of type-A' seller being punished when he mis-sells.

Note that the buyer has full bargaining power in our model. This softens the problem of externality of the existence of a different type of buyers on the buyers we have in the model.

Heterogeneous Sellers: The model assumes that all sellers are immoral in the sense that they mis-sell whenever it is worthwhile for them. Suppose a fraction of sellers are honest, that is, they always truthfully report the appropriate product. It only reduces the buyer's incentive to think, which makes the pooling equilibrium more likely to occur.

In addition, we discuss the case where the seller may not be informed in section 3.3.2.

More than Two States of Nature: Suppose there are not only the appropriate and non-appropriate products, but an order of appropriate products. Simplifying a bit, we have three states of nature: A, A' and A''. After A' is revealed, the buyer may think further to look for the more appropriate A''.

[66] Alternatively, we can reinterpret it as heterogeneous cognitive cost functions. One is C_1 as assumed before. The other is C_2 such that $C_2(0) = 0$ and $C_2(b) = \overline{C}$ for $b > 0$ where $\overline{C}$ is a sufficiently large constant. The buyer with the later function will never think and always contracts with the seller as long as Assumption 3.5 is satisfied. Hence, there is no behavioral difference between these two interpretations, although the beliefs of the buyers in two interpretations are different. Thus, we can model completely unaware buyers "as-if" their cognition costs are significantly high. (See, e.g., Friedman (1953) for the "as-if" justification.)

If the seller only knows A', then our analysis is not modified before the buyer's cognition for A'', because there is asymmetric information only on A' between them. After A' is revealed, we are back to the model by Tirole (2009), since both parties are uninformed about A''. However, if the seller is fully informed, type-A'' seller can pretend to be A or A'. Nevertheless, our qualitative results still hold. That is, when t is small, we still have a pooling equilibrium. Conversely, we have a semi-separating equilibrium where type-A' seller or type-A'' seller (or both) randomizes his choices.

Welfare Comparatives We view social surplus of the buyer and the seller as our welfare criterion. In any outcome of the game, we have only one source for a welfare loss: the cognition cost $C(b)$ for the buyer. Thus, we define the *transaction cost* as the expected cognition cost

$$L \equiv (1-\rho(1-q))C(b)$$

where $\rho(1-q)$ is the probability that type-A' seller opens the buyer's eyes, which is the mere situation in which no cognition efforts of the buyer is involved. Otherwise, the buyer has to bear the cognition cost $C(b)$. In the model, we have three free parameters: ρ, h and t. The comparative statics of the transaction cost with respect to these parameters are in the following proposition.

Proposition 3.9 *If $\rho h \leq C'(1-\frac{t}{h})$, we have $\frac{dL}{d\rho} > 0$, $\frac{dL}{dh} > 0$ and $\frac{dL}{dt} = 0$. Otherwise, we have that $\frac{dL}{d\rho} < 0$, the sign of $\frac{dL}{dh}$ is ambiguous, and $\frac{dL}{dt} < 0$.*

Proof 3.15 *See Appendix A.1.17.*

When $\rho h \leq C'(1-\frac{t}{h})$, there is a pooling equilibrium where type-A' seller mis-sells with certainty. Thus, the buyer exerts her cognition efforts with certainty. The transaction cost therefore is $L = C(b)$. The higher ρ, the higher the cognition level b the buyer exerts in order to avoid being mis-sold to, and thus the higher L. By the same token, L is increasing in h. In the pooling equilibrium, however, the buyer's cognition level b is independent of the transfer t, because changing t does not influence the buyer's marginal payoff of cognition, although the buyer prefers a higher t.

When $\rho h > C'(1-\frac{t}{h})$, there is a semi-separating equilibrium. Since q is endogenous, the welfare comparative statics are not so straightforward as above.

Particularly striking is that we have the opposite result that L is *decreasing* in ρ here. Since, in the semi-separating equilibrium, what determines the cognition of the buyer is only type-A' seller's indifference condition, which is equation (3.31), having more type-A' sellers in the population does not alter the buyer's cognition level. However, when ρ is large, to keep the buyer employing the same cognition level as before, type-A' seller has to reduce q. Equation (3.32) implies q decreases at a higher rate than ρ. Thus, the overall probability that the buyer exerts cognition is lower. Although ρ is higher, there is much higher probability of information disclosure. Hence, L is reduced for a higher ρ.

In a nutshell, *there is a cutoff value $\overline{\rho}$ such that L is increasing in ρ as long as $\rho < \overline{\rho}$ (as in many developed countries) and is decreasing thereafter (as in some less-developed countries).*

In the semi-separating equilibrium, however, whether or not L increases as h increases is ambiguous. Since raising h has an ambiguous impact on equilibrium q, we cannot judge the welfare consequence of it.

Lastly, a higher t reduces L in the semi-separating equilibrium. For a higher t, to guarantee type-A' seller's indifference condition, the buyer's cognition level b is lower. The only way to maintain the buyer's low cognition level is to reduce type-A' seller's probability of mis-selling q. Since both b and q are reduced, L is reduced.

The welfare comparatives suggests that a benevolent court of law should increase t to the largest extent.[67] However, t is bound above by the limited liabilities of the parties (as shown in Appendix A.1.14). Therefore, t can be interpreted as the highest possible transfer, depending on the wealth of the parties in the litigation process.

It is worth mentioning that the results in Proposition 3.9 are robust to the more general case with a direct deadweight loss of mis-selling Δ. To see it, let the transaction cost be $L = (1-\rho(1-q))C(b) + \rho q \Delta$ where $\rho q \Delta$ is the expected welfare loss from mis-selling. It is straightforward to see that $\frac{dL}{d\rho} > 0$, $\frac{dL}{dh} > 0$ and $\frac{dL}{dt} = 0$ in the pooling equilibrium. In the semi-separating equilibrium, since ρq is lower for a higher ρ as shown in Appendix A.1.17, we still have that $\frac{dL}{d\rho} < 0$. Further, the sign of $\frac{dL}{dh}$ is ambiguous as well. Lastly, q is decreasing in t, so $\frac{dL}{dt} < 0$ remains.

Suppose there were a separating equilibrium in which $q = 0$. Then $b = 0$, since it is not worthwhile for the buyer to think any more. Thus $L = 0$, that is, the transaction cost vanishes. However, the separating equilibrium does not exist as shown in Proposition 3.8. Roughly speaking, our ideal society where we have trust everywhere and no wasteful cognition is spent is an impossible world. Nevertheless, we can improve our world by some market institutions as shown in the following sections.

Reputation

In section 4.3.2, we have the negative result of no separating equilibrium, which would be the ideal outcome. However, our discussion so far has focused only on one-shot interaction. We now consider a large economy with a continuum of buyers and sellers

[67] Although, in the pooling equilibrium, t plays no role in the transaction cost, the court of law may choose the highest t in order to enhance the buyer's welfare as a tie-breaking rule.

in repeated relationships.

Assume that time is discrete and infinite. Buyers and sellers are matched randomly in each period. Assume that the state of nature (A is appropriate or not) is drawn randomly in each period, and *independently* across periods. In different periods, different products are traded. Both buyers and sellers are uncertain about the state of nature in the future periods. The buyer does not know the appropriateness of the future product, and the seller does not know which product the buyer will need in the future periods. For example, a buyer who bought a laptop yesterday may buy a mouse (A) today or a status quo printer with high add-on costs of inks (A') tomorrow.

We assume the Poisson death process, i.e., the parties alive at a period remain in the economy in the next period with probability $\lambda \in (0,1)$. Each party who quits at a period is offset by a new party in the period. Let δ_0 be the parties' discount factor. Thus, two parties share the same *relevant* discount factor $\delta \equiv \delta_0 \lambda$. To simplify the exposition, we assume in this section that the transfer is $t = 0$. Ignoring the positive transfer only simplifies the presentation. Thus, the unique equilibrium in the stage game is the pooling one. Since there is no punishment cost for type-A' seller, he mis-sells with certainty in equilibrium in the stage game.

Existence of Separating Equilibrium We now check if a separating equilibrium in each period is possible in equilibrium dynamics.

Consider the following trigger strategies of the parties. Type-A' seller always unveils A', but the buyer pays some price $\widetilde{p} > 0$ *only* after the novel A' is reported, and the buyer does not think at all ($b = 0$). The high price $\widetilde{p}$ can be interpreted as a tip for the seller for bringing the payoff-relevant unforeseen message to the buyer.[68] The actions in the other nodes in the stage game are the same as in Proposition 3.8. However, if some party deviates from it, the new partner can detect this cheating behavior with probability x in each future period. A high x represents a transparent society in which, say, the media can broadly transmit the complaints from some parties, although the contents of those complaints are not verifiable. Detection leads to the outcome that two parties play the pooling equilibrium in this period.[69] The buyer does not pay the tip any more, and type-A' seller does not reveal A'. It is clear that if they play the reciprocal action profile in each period described as above, the transaction cost reduces to $L = 0$.

[68]One may consider the possibility of rewarding the seller in the case where A is indeed appropriate at the post-contractual stage. But, as shown in Appendix A.1.18, the strategy with zero-rewarding when A is appropriate creates the highest possibility for the existence of separating equilibrium in each period.

[69]One may wonder about the possibility of using no trade as the harshest punishment to enhance cooperation between the buyer and the seller in the spirit of Abreu (1986). However, the optimality is not robust to the case with a small probability that the seller is uninformed as well, as we will see in section 3.3.2.

Now we assess the condition under which the trigger strategy profile is sustainable.

First, we check type-A' seller's incentive for a deviation. If he deviates from revealing A', his *net* gain in the current period is $h - \widetilde{p}$. Since the buyer has no pre-contractual cognition, he can successfully mis-sell, but he loses the tip $\widetilde{p}$.

However, the seller has the risk of losing his individual reputation later. The net loss in the future is

$$\frac{\delta\rho}{1-\delta}\left(\widetilde{p} - (1-x)h - x\left(1-b^*\right)h\right)$$

where b^* is the equilibrium cognition level given by equation (3.27). In each future period, if A is appropriate, which happens with probability $1-\rho$, the seller gains zero payoff, irrespective of whether or not he is detected. If A is not appropriate, which happens with probability ρ, the seller loses the tip $\widetilde{p}$ and yet gains the expected rent $(1-x)h + x\left(1-b^*\right)h$ from mis-selling. Here, the seller mis-sells with certainty in each future period by virtue of the stationarity of the game.

Hence, type-A' seller has no incentive to mis-sell if

$$h - \widetilde{p} \leq \frac{\delta\rho}{1-\delta}\left(\widetilde{p} - (1-x)h - x\left(1-b^*\right)h\right),$$

which can be written as,

$$\widetilde{p} \geq \frac{1-\delta+\delta\rho\left(1-xb^*\right)}{1-\delta+\delta\rho}h,$$

which reflects a minimal level of the tip. Thus, the buyer has to commit to remunerate at least $\widetilde{p}^* = \frac{1-\delta+\delta\rho(1-xb^*)}{1-\delta+\delta\rho}h$ to type-A' seller in order to enhance his truthful report. Note that $\widetilde{p}^* < h$, that is, the minimal tip is smaller than the effect of mis-selling. Thus, it may be worthwhile for the buyer to pay the tip.

Second, we check the buyer's incentive for a deviation. If she deviates from paying $\widetilde{p}^*$ for type-A' seller with a truthful report, her net gain in the current period is $\widetilde{p}^*$, where the buyer only pays price zero so as to guarantee the seller's acceptance of the contract in this period.

However, the buyer loses her reputation in expectation, and the net loss in the future is

$$\frac{\delta}{1-\delta}\left(xC(b^*) + \rho x\left(1-b^*\right)h - \rho\widetilde{p}^*\right).$$

In each future period, if the buyer's misbehavior is detected, which occurs with probability x, the two parties play the strategies of the pooling equilibrium in the stage game. The buyer therefore has to bear the cognition cost $C(b^*)$. With probability $\rho x\left(1-b^*\right)$, she is mis-sold to. But she does not have to pay $\widetilde{p}^*$ in the future anyway due to the stationarity of the game.

Hence, the buyer has no incentive to misbehave if

$$\widetilde{p}^{*} \leq \frac{\delta}{1-\delta}\left(xC(b^{*})+\rho x\left(1-b^{*}\right)h-\rho\widetilde{p}^{*}\right). \tag{3.33}$$

Substituting for $\widetilde{p}^{*}$, we conclude that the buyer has no incentive to misbehave if

$$C(b^{*}) \geq \frac{1-\delta\left(1-\rho\left(1-x\right)\right)}{\delta x}h. \tag{3.34}$$

Proposition 3.10 *If the cognition level $C(b^{*})$ in the pooling equilibrium in the stage game is high, both parties are patient (δ is high), the detection probability x is high, and the extent ρ and the effect h of the mis-selling problem are low, then separating equilibrium in each period is likely to exist.*

Since the right hand side of inequation (3.34) is strictly decreasing in δ and x, and converges to zero when δ and x converge to 1, the analysis yields:

Proposition 3.11 *There exist some cutoff values $\overline{\delta}$ and $\overline{x} < 1$ such that for all $\delta \geq \overline{\delta}$ and $x \geq \overline{x}$ separating equilibrium in each period exists.*

In words, if both buyer and seller are sufficiently patient and the probability of detection is sufficiently high, there exists an equilibrium where the mis-selling problem vanishes. In this case, since no cognition is involved, reputation minimizes the transaction cost.[70]

In the equilibrium dynamics, the buyer commits to give the seller a "tip" if and only if the seller brings the surprising news to her before contracting. The tip promotes the seller's incentive to reveal the novel product to the buyer. Since "surprise" is not contractible, however, this tip-institution does not work in the one-shot interaction due to sequential rationality in the stage game. Hence, the convention of tips in a dynamic economy is an instrument of reducing the mis-selling behaviors and pre-contractual cognition.

Persistence of Pre-Contractual Cognition We now use the history-dependence approach initiated by Tirole (1996) to investigate the effect of a one-time shock of mis-selling on the separating equilibrium.

[70]The alternative way of modeling reputation is assuming finite time horizon and yet two types of sellers: an opportunistic one as we assume throughout the paper and a honest one who always reports the appropriate product. (See, e.g., Kreps et al., 1982) In the sequential equilibrium, the opportunistic seller may still truthfully report the product in the beginning periods in order to gain the buyer's trust in the later periods. However, the opportunistic seller must mis-sell at least in the last period. Therefore the buyer has to exert cognitive efforts in the last period. Thus, the assumption of infinite time horizon is crucial for shattering the transaction cost of pre-contractual cognition. But introducing two types of sellers does not alter our results qualitatively.

Suppose, at period 0, A is not appropriate and the economy has a shock such that the seller mis-sells. The thought experiment is as follows. Suppose all sellers truthfully report the appropriate product from period 1 to T (> 0) in order to gain the buyer's trust at period T; and all sellers born at and before period 0 are locked into mis-selling. At period T, if the seller says A is appropriate, the probability that A is indeed appropriate before the buyer's cognition is

$$\begin{aligned} & \frac{1-\rho}{1-\rho+\rho\,(1-x)\,(1-\lambda)\,(\lambda^{T}+\lambda^{T+1}+\lambda^{T+2}+\cdots)} \\ & = \frac{1-\rho}{1-\rho+\rho\,(1-x)\,\lambda^{T}} \end{aligned}$$

which is strictly less than one for all $T < +\infty$. Thus, after period 0, the probability of mis-selling given that A is announced to be appropriate is always positive, no matter how long the sellers truthfully report the appropriate product. Since the buyer's cognition cost function C is continuous, the pre-contractual cognition cannot be completely saved. Whenever mis-selling occurs in some period, for any finite periods of *anti-mis-selling campaigns*, the economy cannot go back to the equilibrium outcome where no pre-contractual cognition of the buyer is exerted. As in Tirole (1996), however, the out-of-equilibrium *amnesty*, which destroys the evidence of this mis-selling spot, can let the economy return immediately to the separating equilibrium outcome where the transaction cost vanishes. In other words, the "big-bang" approach is more effective than the gradualist strategy in the context of mis-selling problems. We summarize the results in the following proposition:

Proposition 3.12 *The policy of anti-mis-selling campaigns cannot shatter pre-contractual cognition, but amnesty saves the pre-contractual cognition completely.*

Secret Awareness of the Seller The analysis in the last subsection focuses on the effect of a one-time *generation-specific* shock of mis-selling. Here we consider the possibility that mis-selling may come from an infinite cost of acquiring the product-appropriateness information for some seller in some period (whereas we assumed that the seller is always informed about the product appropriateness, or equivalently the cost of acquiring the information is zero in the baseline model.)

Based on Green and Porter (1984), we assume that when A is not appropriate, in each period, the seller is as unaware as the buyer with probability α. The buyer cannot observe whether or not the seller is aware. Now it is not optimal for the buyer to punish the seller forever in the case of mis-selling, because mis-selling may be unavoidable in some situations, and is therefore not the seller's liability. Suppose that the seller who sells an inappropriate product in period 0 is punished from period 1 to T whenever the seller is detected in these periods.

Let V^+ (respectively, V^-) denote the present discounted value of the seller's payoff beginning from a cooperative phase (respectively, non-cooperative phase). Formally, we define them recursively:

$$V^+ = (1-\alpha)\left(\rho\widetilde{p} + \delta V^+\right) + \alpha\left(\rho h + \delta V^-\right)$$

and

$$V^- = \frac{1-\delta^T}{1-\delta}\rho\theta h + \delta^T V^+$$

where $\theta \equiv (1-x) + x(1-b^*) \in (0,1)$ is probability of successful mis-selling given that A is not appropriate in a period from period 1 to T. Starting from the cooperative phase, with probability $1-\alpha$, the seller receives the expected tip $\rho\widetilde{p}$, and the two parties continue to cooperate in the next period. With probability α, the seller is uninformed, and therefore gains the expected rent ρh by chance. Then they switch to the non-cooperative phase. In the beginning T periods of the non-cooperative phase, they play the pooling equilibrium in the stage game where the seller receives the expected rent $\rho\theta h$. Thus, the inefficient cognition occurs in these T periods. After these T periods, they switch to the cooperative phase.

On the seller's side, we have to consider the incentive constraint of the informed type-A' seller. The aware seller will not pretend to be unaware:

$$\widetilde{p} + \delta V^+ \geq h + \delta V^-$$

which can be written as,

$$\widetilde{p} \geq \widetilde{p}^*(T) \equiv \frac{1-\delta+\delta\left(1-\delta^T\right)\left((\theta-\alpha)\rho+\alpha\right)}{1-\delta+\delta\left(1-\delta^T\right)\left((1-\alpha)\rho+\alpha\right)}h, \tag{3.35}$$

which implies that

$$\frac{d\widetilde{p}^*(T)}{dT} = \frac{\rho\delta^{T+1}(1-\theta)(1-\delta)\ln\delta}{\left(1-\delta+\delta\left(1-\delta^T\right)\left((1-\alpha)\rho+\alpha\right)\right)^2}h < 0.$$

Proposition 3.13 *The minimal tip $\widetilde{p}^*(T)$ that the buyer has to reward the seller for not deviating is decreasing in the length of punishing the seller T.*

Intuitively, the tip $\widetilde{p}$ as a carrot and punishment length T as a stick are substitutes for the buyer to incentivize the seller's cooperative behavior.

Since mis-selling may occur by chance, our social goal is to minimize the punishment periods T in which the cognition costs of the buyer are involved. However, reducing T raises $\widetilde{p}^*(T)$. On the buyer's side, the buyer's incentive constraint for not deviating is the same as before. Thus, $\widetilde{p}^*(T)$ is bound above by the buyer's incentive constraint (3.33). Inequation (3.33) implies $\widetilde{p}$ cannot exceed

$$\frac{\delta x \left(\rho h \left(1 - b^*\right) + C\left(b^*\right)\right)}{1 - \delta + \rho\delta}.$$

Plugging it into equation (3.35), we have that in the equilibrium dynamics, the smallest punishment periods T^* has the property that $\frac{dT^*}{d\alpha} > 0$. Therefore, we obtain the following proposition:

Proposition 3.14 *In the equilibrium dynamics, the pre-contractual cognition is strictly increasing in α which is the probability that type-A' seller is unaware of A'.*

Hence, the lack of knowledge of the seller that is unavoidable raises the transaction cost. In other words, specialization of the seller reduces the transaction cost.

Competition

In this section, we extend the model in section 4.3.2 to a setting with two sellers (S_1 and S_2) who simultaneously report the product in stage 2. The two sellers here are in the same industry. Thus, the product-appropriateness is industry-specific. The two sellers belong to the same type. If A is appropriate, they can only report A, but if A' is appropriate, the payoff matrix of their interaction situation is described in Table 3.1.

S_1, S_2	A	A'
A	$(1-b^*)(\frac{h}{2}-t)+b^*(-t)$, $(1-b^*)(\frac{h}{2}-t)+b^*(-t)$	$-t$, 0
A'	0, $-t$	0, 0

Table 3.1: Payoff Matrix of Competition for two type-A' Sellers

In the payoff matrix, b^* is the buyer's equilibrium cognition level in this competition context. If both sellers choose A', they obtain payoff zero, no matter whose product is chosen by the buyer in the end, which is due to the buyer's full bargaining power. If, say, S_1 chooses A and S_2 chooses A', the buyer is immediately aware of A'. Then the buyer sues S_1 for a transfer t. But S_2 gets zero payoff. If they both choose A, they obtain their expected payoff. We assume that the two sellers have equal probability of being chosen. The gain of mis-selling is therefore $\frac{h}{2} - t$, since only one of their products can be chosen by the buyer, and the buyer can sue both sellers at the post-contractual stage.

In the case of competition, we have the following proposition.

Proposition 3.15 *Suppose A is not appropriate. If $\frac{h}{2}-t<0$ or $\rho h>C'(1-\frac{2t}{h})$, truth-telling by two sellers (A', A') is the unique equilibrium. Otherwise, we have multiple equilibria: truth-telling (A', A'), mis-selling (A, A) and the totally-mixed-strategy one. In the latter case, the disclosure decisions of A' are strategically complementary for the two sellers.*

Proof 3.16 *See Appendix A.1.19.*

As one would expect, truth-telling as an action profile (A', A') by two type-A' sellers is an equilibrium in the Nash implementation fashion (Maskin, 1999). Given that one seller unveils A', the other seller has no incentive to mis-sell due to the punishment t for mis-selling. Furthermore, if $\frac{h}{2}-t<0$, announcing A' is also a strictly dominant strategy. In this case, the equilibrium (A', A') is therefore unique. Competition reduces the rent of mis-selling for the seller, and yet does not alter his cost of the transfer. Thus, no seller has an incentive to mis-sell.[71] Therefore, when competition is present, there exists separating equilibrium in the sense that both types of sellers truthfully report the appropriate product, which reduces the transaction cost to zero.

Nevertheless, competition is by no means a panacea. When $\frac{h}{2}-t\geq 0$ and $\rho h\leq C'(1-\frac{2t}{h})$, there are multiple equilibria. In essence, two type-A' sellers' disclosure decisions of A' are strategically complementary. That is, they reveal their private information hand in hand here.[72]

To sum up, competition *may* induce separating equilibrium, which minimizes the transaction cost. This result is in contrast to Tirole (2009). In the context of asymmetric information and the legal punishment for the seller who mis-sells, competition between sellers plays a role of reducing the transaction cost.

Concluding Remarks

Policy Recommendations Atkinson et al. (2000) suggest *ex ante* solutions to the mis-selling problem mainly via promoting public awareness, say providing independent advice to consumers, and educating consumers through mass media, schools and so on. Korobkin (2003) also recommends *ex ante* intervention by legislatures, say mandatory

[71]It implies here that the number of sellers and the transfer t are substitutes. Hence, introducing a sufficient number of sellers can always achieve the unique equilibrium (A', A') even though t is bounded above. In reality, however, only a limited number of sellers enter the market due to some entry barrier. Thus, without loss of generality, we assume there are only two such sellers.

[72]Notice that we can design the competition mechanism differently to let equilibrium (A', A') more likely to be unique. If, say, S_1 announces A and S_2 announces A'; we transfer the punishment t for S_1 to S_2, then a weaker condition $\frac{h}{4}-t<0$ guarantees the equilibrium uniqueness.

information disclosure rule. However, the *ex ante* mechanism is valid only if the regulator knows the product-appropriateness in each industry, which is not plausible.

One may consider the possibility of using catch-all clauses in the law. For example, the legislator can require the seller to disclose all the add-ons without the need of describing the particular names of the add-ons. However, catch-all clauses are always vague. For instance, besides the cost of inks, consumers also need printing paper, a computer, and even electricity for using a printer. It is not reasonable, however, to define these additional costs as add-ons. Moreover, how the seller discloses the eye-opening information is relevant. If the seller puts the information only in fine print, it is nothing but mis-selling in our model. The cognition cost of the buyer turns to be the cost of reading the fine print. Further, fine print also weakens the instrument of warranties to solve the mis-selling problem, because the fine print per se involves a new problem of the buyer's limited cognition and the seller's "mis-presentation".

Some scholars suggest the *ex post* judicial mechanism, say by using the Unconscionability doctrine to interpret contracts, which refuses to enforce those contracts with unconscionable terms.[73] However, this doctrine as applied by common law courts is not defined by status, and thus is too vague. By contrast, we suggest a mechanism in Appendix A.1.14 that may reduce the transaction cost. However, this judicial mechanism is still imperfect.

Nonetheless, our analysis shows that, without an omniscient legislator and a perfect judicial system, we can still achieve the ideal separating equilibrium in which the transaction cost vanishes via the market institution, namely reputation or competition.

First, to ensure the separating equilibrium by the reputation mechanism, we advocate "tips" in the economy, which require that both buyer and seller are sufficiently patient and the probability of detecting the past cheating behavior is sufficiently high, that is, x and δ are high enough. A high δ can be realized through some effort to reduce the mortality rate $1-\lambda$. A high x can be achieved through freedom of the press, which means journalists have the opportunity to move without friction and to report scandals without any interference. In the case where it is commonly known that the seller is well-informed about the product-appropriateness, if the out-of-equilibrium shock of mis-selling occurs, amnesty can let the economy go back to the outcome where the pre-contractual cognition is completely saved. Thus, a firm with a mis-selling spot in the past should change its brand. Lastly, if the seller may be unaware as well, we should promote the specialization of the sellers in any industry to increase the probability of the seller's awareness α in order to reduce the transaction cost.

Second, to promote voluntarily information disclosure by the seller of buyer-unforeseen features of the product, we should reduce the entry barrier in the industry so as to

[73] See Korobkin (2003) and Becher (2008).

induce competition among sellers.

It is worth noting that in the separating equilibrium neither mis-selling nor litigation occurs. Thus, even in a world where the transaction cost from direct mis-selling and litigation are non-negligible, these two market institutions still achieve the first best outcome.

Applications The analytic framework of the seller's strategic mis-selling against the buyer's pre-contractual cognition in the case of asymmetric awareness can be applied to numerous problems.

In financial markets, investors may be uncertain as to whether or not the firm has false statements, but they can employ some accountant to audit the firm. Audit here can be interpreted as cognition. In contrast to the conventional audit literature, if the accountant finds that the firm is cheating, it must provide hard evidence. Otherwise, the investors remain uncertain, since finding nothing provides only soft information. In addition, the firm can also provide hard evidence of its own flaw. It would be interesting to study the strategic statements of the firm in the accounting problems.

In politics, there is also asymmetric information on the appropriateness of platforms between the politicians and voters. It would be also interesting to study strategic disclosure of flaws in competitors' platforms in political competition, given that the voters can make efforts to contemplate on the platforms.

Hence, our model is merely a starting point. These further issues give us a rich outline for the future research in asymmetric awareness and pre-contractual cognition in broader social problems.

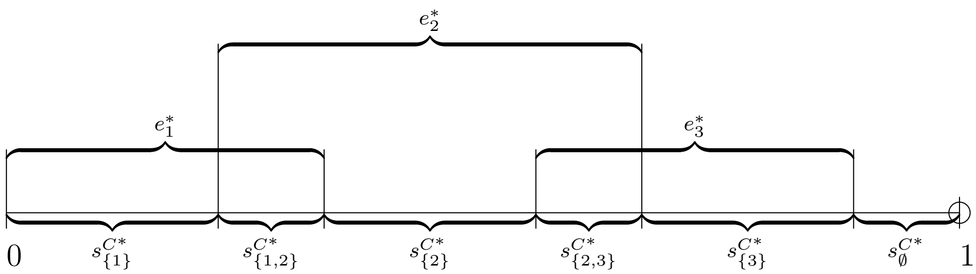


Figure 3.5: An equilibrium contract C^* when $K = 3$

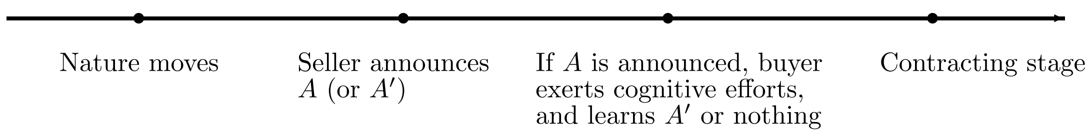


Figure 3.6: Timeline

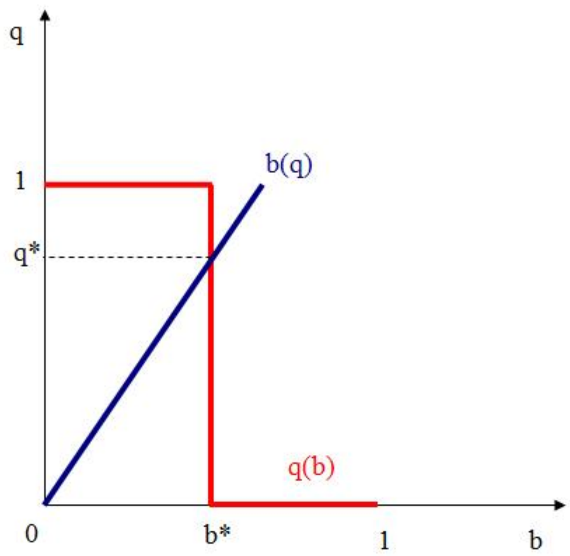


Figure 3.7: Semi-Separating Equilibrium

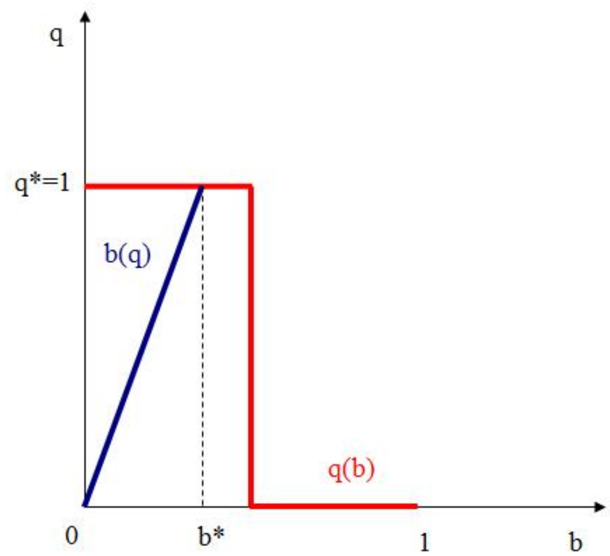


Figure 3.8: Pooling Equilibrium

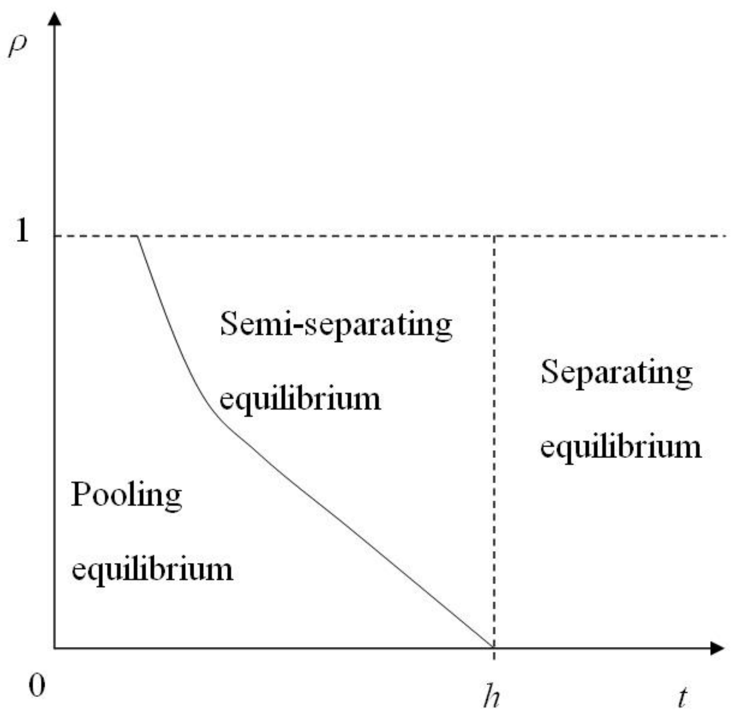


Figure 3.9: Graphical Illustration of Proposition 3.8

Chapter 4

Unawareness of Actions

"The way grew by one foot but the demon grew by ten."

Wu, Cheng'en *Journey to the West* (1590, chapter 50)

Unawareness is a common aspect of ignorance of human being in economic life. At micro economic level, many unforeseen events are directly man-made, meaning, an agent's surprise is as a result of the actions of the others. An investor may be unaware that an entrepreneur takes the investment to casinos. An insuree may be unaware that an insurer delays the payment for his life insurance. A car-buyer in the second hand market may be unaware that a seller sells a car, which is not legally owned by this seller. Most printer-buyers are unaware that their later costs of the ink are very high. (Gabaix and Laibson, 2006)

In particular, this kind of phenomena worsens the "moral hazard" problem in the Principal-Agent relationship. Consider an employee who has accepted a contract proposed by an employer. In the post contractual stage, the employee may regret accepting it, because the quality of the working environment is terrible. Since the employee was unaware of this utility-relevant dimension when signing the contract, he has to bear this bad working condition ex post.

The standard moral hazard model in contract theory implicitly assumes that the contracting parties (the principal and the agent) are fully rational. Although the principal cannot observe the action of the agent, he knows that he cannot observe it and the entire action set of the agent is in the principal's mind. Hence, the principal can design the incentive scheme given the agent's incentive response. However, the principal's problem is not only confined to unobservability. It is also possible that some action of the agent is unimaginable to the principal. For example, some inexperienced employer may be unaware that his employees could manipulate the short run working performance at the expense of the long run benefit. On the other hand, the principal could also do something out of the agent's mind without violating the contract. Moreover, the agent may also be unaware of his own choice possibilities, and the principal has an incorrect belief of the agent's awareness.

When unawareness is present in the contracting situation, what occurs when two parties interact? Is awareness of a party always valuable for him? Zhao (2008) addresses these questions.

To answer the first question, the solution concept of the standard moral hazard model is not satisfactory, because it assumes that both parties are fully aware and thus are able to optimize. Based on information structures with unawareness, the paper extends the standard model to a model with unawareness of *actions*. The primary goal of the paper is to provide a generalized solution concept. The contracting parties involved are still rational utility-maximizers. However, they are only "locally" rational. In terms of Arthur (1994), they make logical deductions based on their current hypotheses of the situation. The principal designs the contract according to his mental model of the contractual setting. Moreover, he contemplates the incentive scheme according to the his conjecture of the agent's subjective contractual setting, and decides how to update the agent's awareness through the contract. In the post contractual stage, either party may be surprised by his opponent's action.

Concerning the second question, we show that a party who is more aware of his own action sets can be strictly worse off. The underlying intuition is that a party may still be harmfully surprised by some of his opponent's actions, which were out of his mind, even though he is more aware of his own action sets. A realistic example is that a firm (principal) contracts with a research institute (agent) to invent some new technology to improve the firm's productivity. Suppose the firm is additionally aware of the possibility that it can sell the machines with the new technology to some other firms. The firm sells the machines and believes that it can earn a higher profit. However, the firm is unaware that the research institute obtains a patent on the new technology. Thus the firm suffers from its illegal behavior of selling the machines ex post. Awareness of selling the machines makes the firm strictly worse off. However, under certain conditions, which are characterized in the paper, increasing awareness of a party's action sets weakly increases his utility.

Moreover, under the condition that the agent is not more aware than the principal believes, if the principal is more aware of the agent's action sets, he is weakly better off. The intuition is that if the principal is aware of some additional action sets of the agent, which are out of the agent's mind, then the principal could strategically choose whether to inform the agent about the additional action sets for his own interest. For example, if an employer is additionally aware that it is more efficient to let his employee to get some external skill training to increase the productivity, he will inform the unaware agent of this possibility through the contract. The additional awareness makes the employer better off.

Finally, we show that if the agent is more aware of the principal's action sets, the agent is not necessarily weakly better off, *even if the agent becomes aware of all action sets of which the principal is aware.* The reason is that unawareness of the agent guarantees that the agent chooses some certain action. However, after becoming more aware, the agent will deviate from this certain action. Taking it into account, the principal will propose a different contract. Thus the agent cannot achieve the former utility anymore.

4.1 Contractual Traps

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 that 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 not reasonable, 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 work for him. 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. This unawareness is modeled by introducing the missing dimensions of the strategy set, which is akin to the unawareness models of missing dimensions of the full state space (see, e.g., Li (2009)). 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.

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 solution* 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 of the game. This discrepancy creates room for various choices of alternative solution concepts, as we elaborate below.

The second solution concept we introduce is the *justifiable solution.* Under this solution concept, 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 agent's reasoning upon receiving a contract alters what the principal is able to offer, thereby giving rise to an additional "justifiability" constraint on the principal's side. The justifiable solution is intended to capture the idea that an unaware agent may still be able to evaluate whether the principal's contract offer is "reasonable" (see Filiz (2008), Ozbay (20089, and more fundamentally Heifetz et al. (2009)). It is also 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 (Kohlberg, 1986).

In the third solution concept, we intend to capture the idea that while confronted with an unintended (non-justifiable) contract, the agent may believe that this unintended contract simply results from *the principal's mistake* occasionally.[1] In such a scenario, we can conveniently assume that from the agent's perspective, a non-justifiable contract results from the principal's mistake with probability $1 - \rho$, 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 solution*. 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 solution degenerates to a rational solution. On the other hand, if $\rho = 1$, whenever she sees an unintended contract, she perceives it as a trap and the trap-filtered solution coincides with the justifiable solution. Thus, the trap-filtered solution 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 (2009), 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 solution with cognition*, the agent does not exert cognitive effort only if she sees a justifiable contract. In contrast, in Tirole (2009), the agent will not exert cognitive effort only if the principal opens the agent's eyes.[2]

Our main contribution is to provide a general framework that unifies a number of seemingly unrelated solution concepts in Filiz (2008), Ozbay (2008), Tirole (2009), von

[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., McKelvey and Palfrey (1995)

[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 4.1.2 for details.

Thadden and Zhao (2009) and Zhao (2008), which allows us to investigate the interactions among unawareness, reasoning, and cognitive thinking in the optimal contract design context. This framework allows us to gain a deep understanding of the economic agents' decision making while potentially confronted with a contractual trap; additionally, through investigating the agent's response, the firm (as the principal) can better design their contractual terms based on the managerial implications generated in this paper. In Section 4.1.3, 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 sophistication is crucial.

Since we incorporate unawareness to the principal-agent relationship, our paper is related to vast literature on the unawareness. Fagin and Halpern (1988), Modica and Rustichini(1994) and Modica and Rustichini(1999) first discuss the unawareness issue formally, and Dekel et al. (1998) 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 (2009) 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 Filiz (2008), Gabaix (2006), and von Thadden and Zhao (2009). 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 More (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 (2009)). 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

(2009) introduces the contract incompleteness from a very different angle. Namely, in Tirole (2009), 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 4.1.1, we introduce the principal-agent framework, and in Section 4.1.2 we propose a number of solution concepts and discuss the behaviors under those solution concepts. In Section 4.1.3, we demonstrate the implications of these solution concepts in an example. Section 4.1.4 concludes.

4.1.1 The Model

We consider a stylized model in which a principal (P) intends to hire an agent (A) to work for him and 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 decisions in many dimensions, here $S_P \equiv A_P^1 \times \ldots \times A_P^M$ 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 employee, 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. The set S_A includes these decisions of the employee.

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 set of strategy profiles, 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 to a real-valued utility. Notably, the general formulation here has incorporated uncertainty into the utility functions.[4] If eventually the agent rejects the contract, they

[3] In this way, the games (to be defined later) will be finite as well. The existence of solution (under the appropriate solution concepts) can be easily established following the classical game theory literature (see, e.g., Fudenberg and Tirole (1991)).

[4] For example, in the classic employee-compensation 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 contingencies, the probability distribution of the residual uncertainty ϵ should be common knowledge.

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. Along the line of the modeling technique initiated by Li (2009), 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 set of strategy profiles S and the agent's awareness (i.e., W).[5] 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.[6]

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 actions that the agent is aware of (i.e., W); additionally, the principal might announce actions that are out of the agent's mind. We use $V \equiv V_P \cup V_A$ to represent the collection of action sets that are specified in the contract but are out of the agent's mind. The set V can be interpreted 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 the notation $\times X$ to denote the Cartesian product of all action sets in $X \subseteq D$, i.e., $\times X \equiv \Pi_{Y \in X} Y$.

A **contract** is a vector $\psi(V) \in \times(W \cup V)$ where $V \subseteq D \setminus W$.[7] Note that $\psi(V)$ specifies all actions that the agent is aware of after observing the contract. 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 (2008) and

[5] The situation where the principal is uncertain about the agent's awareness and therefore screens the agent's awareness is studied by von Thadden and Zhao (2009).

[6] 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.

[7] 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 $\times W \equiv A_P^1 \times A_P^3 \times A_A^2$.

Ozbay (2008).

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

Definition 4.1 *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 *welfare-relevant* actions that party i can select.[8] 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.

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. Likewise, for $A_P^k \notin W_P \cup V_P$, the agent unconsciously assumes that the principal will choose the default action $\bar{a}_P^k$. Since it is an unconscious choice, it is natural to assume that the default action is unique. In other words, the contracting parties have no disagreement about the "unconscious" default actions. For example, if the agent may be unaware of playing "Second Life" in her office, the default action in this dimension is simply not playing it. If the agent may be unaware that the principal can delay the salary payments, the default action of the principal is not to delay them.

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*

[8]It is worth mentioning that Tirole (2009) 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 utility-relevant actions.

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 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 an unconscious hypothesis in the agent's mind. The agent is unaware that 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 action profile that the agent is unaware of, where the superscript C stands for "complement." 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 $\overline{a}_P^k$. 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 the agent's unawareness and rule-guided behavior, 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

[9] The term "subjective" means subjectivity of the agent's belief but not subjectivity of her preference. A belief-subjective utility could be wrong, since the agent "believes" that the utility has certain form that is a hypothesis in his mind. On the other hand, we cannot argue whether a preference-subjective

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

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

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 $u_i^V(\cdot)$ are coherent 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.[10] Since in our context the agent updates/expands the subjective utility function to her objective utility, our notion follows the modeling strategy dating back to Modica (1998), where they study a general equilibrium framework.

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 agent accepting the contract 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.

utility is wrong or not, because the utility reflects the agent's true "feeling" that represents her personal value judgment.

[10]An alternative way to model the set of strategy profiles is to define a correspondence $M : 2^S \mapsto 2^S$ from an announced subset of S, denoted by Y, to the updated action sets $M(Y)$ in the agent's mind after the principal's announcement. Note that $M(Y) = M_P(Y) \times M_A(Y)$ specifies both the principal's and the agent's strategy sets. By this formulation, a contract $\varphi = (\varphi_P, \varphi_A)$ is an element in $M(Y)$.

This alternative model sounds more general and flexible. However, it is not convenient to model how the principal deviates his specified actions φ_P in the contract in a natural way. In fact, this deviation plays an important role in the problem of contractual traps. On the contrary, our modeling framework avoids this difficulty since the principal can freely choose any actions in the dimensions out of the agent's mind, whereas the principal has to fulfill the actions in the dimensions in the contract, which the agent is aware of.

[11]This may not be appropriate in certain scenarios, but modifications are straightforward. On the one hand, if all actions are verifiable, the strategy of the agent can be directly written into the contract, and thus unawareness of agent does not matter. On the other hand, if no action is verifiable, when 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.

4.1.2 Solution Concepts

In this section, we provide predictions of the 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 the terminology of game theory, these rules are described by the "*solution concepts*" (see Fudenberg and Tirole (1991)). 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, and then introduce a number of solution concepts that are suitable for the economic environments that involve unawareness.

Preliminaries

Before introducing the solution concepts, we specify the timing in this contractual relationship 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 a *bundle*. Given the contract and the agent's updated awareness, we can describe how the agent chooses her strategy and whether to accept the contract or not. As in the standard principal-agent problems, the choice of the agent in the contract 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 of a specific strategy profile $\widetilde{\psi}_A(V)$ and $\psi_P(V)$ in the agent's mind. The incentive compatibility constraint guarantees that the strategy in the contract 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 4.2 *A contract $\psi(V)$ is **feasible** if it satisfies (IC) and (IR).*

The above discussions concern whether the agent is willing to follow the proposed behavior.

Definition 4.3 *A bundle $(\psi(V), s)$ is **coherent** if $\psi(V) = s(V)$ and $s_A^C(V) = \overline{s}_A(V)$.*

The coherence of a bundle ensures that the principal's realized actions are the same as his claimed actions in the contract and the agent chooses the default actions in the dimensions she is not aware of. Feasibility and coherence are maintained throughout this paper in every solution concept, as we describe next.

Rational Solution

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

Definition 4.4 *A bundle $(\psi^*(V^*), s^*)$ is a **rational solution** if the principal chooses V^*, $\psi^*(V^*)$, and s^* that maximize u_P s.t. $\psi(V)$ is feasible and $(\psi(V), s)$ is coherent.*

To interpret the rational solution, 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. Thus the information conveyed in the contract must be optimal from the principal's perspective. Note that since we have assumed that $\inf_{s\in S} u_P(s) \geq \overline{u}_P$, the principal strictly prefers to have the agent participate. Further, as A_i^k is finite for all i and k, the rational solution exists because the game is finite as well (see, e.g., Fudenberg and Tirole (1991)).[12]

The rational solution 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, a player simply ignores how she reaches the node. All what matters is the future. Due to this subgame perfect feature, the game is solved by *backward induction.* As we apply the subgame perfect Nash equilibrium to our context, we shall first focus on the agent's problem. Here, the novel

[12]It is worth mentioning that the existence of the rational solution can be guaranteed here, since we implicitly assume that there is no probabilistic announcement of contracts by the principal.

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 of the game. This discrepancy creates room for various choices of alternative solution concepts, as we elaborate in the subsequent sections.

Justifiable Solution

In the rational solution, 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 assumed in Filiz (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 agent's viewpoint) among all the feasible contracts. This gives rise to the next solution concept, namely the justifiable solution.

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

Definition 4.5 *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 coherent, 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 something may go wrong.[13] This assumption is also adopted in Filiz (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 solution). All the above descriptions

[13]Alternative models of awareness of unawareness can be checked in Halpern and Rego (2006), Rego and Halpern (2007) and Tirole (2009).

require a higher-order reasoning of the agent. 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 offers a contract that is not reasonable, 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 $\inf_{s \in S} u_A(s) < \overline{u}_A$, 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 a "unreasonable" contract offer, it is convenient to assume that in the agent's mind, a contractual trap leads to the worst-case utility $\inf_{s \in S} u_A(s)$.[14]

One may argue that adopting the lowest utility level here is a special case of ambiguity aversion. In fact, as long as we assume that the agent's perceived utility Z_A from a non-justifiable contract is worse than her outside option, the agent will reject the contract anyway. Thus, the crucial assumption we make here is that an agent who is pessimistic about a non-justifiable contract. Alternatively, one may consider the possibility that the agent can derive the worst possible outcome within her awareness, i.e., $Z_A = \inf_{s_A \in W_A \cup V_A} u_A^V(s\,(V))$. This alternative scenario essentially makes no difference if the derived worst outcome within the agent's awareness is also worse than her outside option.[15]

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

[14] The assumption that the agent knows her worst utility even if she is unaware can be justified by the limited liability of the agent. For example, the worst outcome for an investor is usually known: zero return.

[15] Note that this alternative scenario has its own issue. Facing a non-justifiable contract, the agent is aware that something may go wrong and therefore she knows that her awareness is limited. Thus, it is no longer plausible that the agent still employs the derived worst outcome and uses this to compare with her outside option.

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 solution.

Definition 4.6 *A bundle $(\psi^*(V^*), s^*)$ is a* ***justifiable solution*** *if the principal chooses V^*, $\psi^*(V^*)$, and s^* that maximize u_P s.t. $\psi(V)$ is justifiable and $(\psi(V), s)$ is coherent.*

In a justifiable solution, 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 solutions, 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. The existence of a justifiable solution can be easily established.

Lemma 4.1 *There exists a justifiable solution.*

Proof. To prove the existence, first observe that there exists at least one justifiable contract: the one that makes the agent fully aware, albeit it may be suboptimal. Now if the principal chooses the optimal contract among those feasible contracts that are justifiable for him as well, we then obtain a justifiable solution according to the definition. □

The idea of justifiable solution 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 (1991)). In 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 alters what the principal is able to offer. Moreover, this solution concept 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 a non-justifiable contract follows from our assumption that $\inf\limits_{s \in S} u_A(s) < \overline{u}_A$.[16]

It is worth mentioning 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

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

justifiability of the contract. The agent has a wrong belief regarding the principal's sophistication off the equilibrium path, although on the equilibrium path the agent's belief is correct. See also Ozbay (2008) for the discussions on justifiability in a different context. Our notion of justifiability is in the same spirit of the cognitive hierarchy (the generalized level-k thinking) discussed in Camerer et al. (2004): by imposing the consistent belief only on the equilibrium path, we assume that the agent simply adopts the level-1 thinking. It is possible to extend our analysis to higher-level cognitive thinking, but such an extension necessarily complicates the presentation of the solution concepts.[17]

So far we have introduced two different solution concepts. In a rational solution, the agent takes the contract as given and updates her awareness passively. On the contrary, in a justifiable solution, 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 unifies two extremes. This motivates us to propose the next solution concept.

Trap-filtered Solution

In the justifiable solution, 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 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 McKelvey and Palfrey (1995).

Our goal in this section is to incorporate this type of bounded rationality into the unawareness framework. Formally, when the agent faces a contract $\psi(V)$ that is not

[17] Notably, we may also allow justifiability on the off-equilibrium paths following the argument of dynamic unawareness and rationalizable behaviors (see, e.g., Heifetz et al. (2009)) that extends the classic game theory literature, including extensive-form rationalizability (Pearce, 1984). Although some algorithmic procedure makes the solution easy to compute via computer programs, it involves a very high order of interactive reasoning of the contracting parties. Since the aim of our paper is to analyze contractual traps economically yet not "computer" traps mechanically, we adopt more realistic solution concepts here.

justifiable, 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.[18] In the appendix, we illustrates the determination of ρ in details. With these probabilities, we can then express the agent's expected utility upon observing a non-justifiable contract $\psi(V)$ as follows:

$$U_A^T(\rho, \psi(V)) \equiv \rho Z_A + (1-\rho) u_A^V(\psi(V)),$$

where Z_A corresponds to the utility the agent attaches to herself 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.

As discussed before, this utility Z_A can be set as some exogenous utility level the agent believes to obtain after observing a non-justifiable contract. As before, a convenient way to assign a value to Z_A is based on the recent advances on "ambiguity aversion," in which case $Z_A = \inf_{s \in S} u_A(s)$ corresponds to the agent's worst-case utility. This extremely pessimistic perception follows from the intrinsic ambiguity aversion the agent may exhibit, and naturally there are other good candidates for Z_A that are less pessimistic. Nevertheless, to fix ideas, in the sequel we choose $Z_A = \inf_{s \in S} u_A(s)$ for ease of exposition.

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 4.7 *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 (IR-T) need not be justifiable.

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

Definition 4.8 *A bundle $(\psi^*(V^*), s^*)$ is a **trap-filtered solution** if the principal chooses V^*, $\psi^*(V^*)$ and s^* that maximize u_P s.t. $\psi(V)$ is trap-filtered and $(\psi(V), s)$ is coherent.*

[18] It potentially bridges the gap between the two orthogonal solution concepts – the forward induction and the trembling hand equilibrium – in game theory. The conceptual difference between the two solution concepts has been well documented, see, e.g., van Damme (1989).

In the appendix, we rationalize this solution concept through a game with a lexicographic probabilistic system by Blume et al. (1991). The existence of this solution concept is guaranteed by a straightforward proof that naturally extends the proof of the existence of justifiable equilibrium.

Note that when $\rho = 0$, the agent is extremely confident that any non-justifiable 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 solution degenerates to a rational solution. On the other hand, if $\rho = 1$, the agent believes that the principal never makes a mistake; thus, whenever she sees a non-justifiable contract, she perceives it as a trap and the trap-filtered solution coincides with the justifiable solution. Thus, the trap-filtered solution can be regarded as a broader family of the solution concepts that incorporate the ones reported in the literature. The existence of solution follows the similar arguments and therefore is omitted.

Trap-filtered Solution with Cognition

The trap-filtered solution 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 discuss in the appendix; 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 (2009), 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 (after the contract is announced), 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.[19] 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 $\rho(1-c)$, the agent cannot figure out the trap and attaches ex ante the utility $\inf\limits_{s\in S} u_A(s)$ 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:

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

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:

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

[19] This is different from Tirole (2009) in that cognitive thinking occurs when the contract is not justifiable, whereas in Tirole (2009), 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 (2009).

Note also that cognitive thinking allows the agent to determine whether to accept the contract *after* the cognition stage.[20]

We can now introduce the solution concept with cognition.

Definition 4.9 *A contract $\psi(V)$ is* ***trap-filtered with cognition*** *if it is justifiable, or it is feasible and (IR-C) holds.*

Definition 4.10 *A bundle $(\psi^*(V^*), s^*, c^*)$ is a* ***trap-filtered solution with cognition*** *if the principal chooses V^*, $\psi^*(V^*)$ and s^* that maximize $\{c^*\overline{u}_p + (1-c^*)u_P(s)\}$ s.t. $\psi(V)$ is trap-filtered with cognition, $(\psi(V), s)$ is coherent, and $c^* = 0$ if $\psi(V)$ is justifiable and $c^* \in \arg\max_{c' \in [0,1]} U_A^C(c', \rho, \psi(V))$ otherwise.*

Along the lines of the appendix, 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 solution with cognition, the rational principal intends to choose $V^*, \psi^*(V^*)$, and s^* that 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 $\rho(1-c)$).

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. Nevertheless, a finite game is sufficient for existence but not necessary. As we demonstrate in Section 4.1.3, a trap-filtered solution with cognition may still exist even if the strategy space is not finite. In general, if T is weakly convex and continuous, then existence of trap-filtered solution with cognition can be established following the arguments in Debreu (1952).

[20]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.

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 one example to demonstrate the similarities and differences of these solution concepts.

4.1.3 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 $\{\overline{a}_i^1\}$ 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, $W = \{A_P^1, A_A^1\}$ 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.[21] We 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.[22] 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

[21]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.

[22]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.

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, $V = \{A_A^2\}$), 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 solution, 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 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 solution. The principal proposes the novel car for the agent, but does not mention the possibility of removing the air conditioning. Notably, this solution concept 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 solution 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 solution 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 solution 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 $c = \rho$, 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 solution 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 solution 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.

4.1.4 Concluding Remarks

In this paper, we provide a general contracting framework to investigate the strategic interactions with the unawareness, reasoning, and cognition, and propose several solution concepts in various degrees of the agent's sophistication. 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 (2009), 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.

4.2 Unaware Consumers and Exploitations

Gabaix and Laibson (2006) consider shrouded attributes in product contracts. In the market, there are many buyers (B) and sellers (S). The seller i produces the product with zero cost and chooses the price of a base good p_i and the price for an add-on $\widehat{p}_i \in [0, \overline{p}]$. Additionally, the seller can choose to shroud the information about the add-on or not. There are two kinds of buyers: the buyers with fraction λ who are aware of the add-on and the buyers who are completely unaware of the add-on. Moreover, the buyer can exert effort cost e to substitute away from future use of the add-on ex post.

The aware buyer's net surplus from choosing seller i is

$$x_i \equiv -p_i - \min\{E\widehat{p}_i, e\} - (-p^* - \min\{E[\widehat{p}^*], e\})$$

where * represents the best alternative seller. The aware buyer forms a rational expectation about the seller's add-on price.

The unaware buyer considers only the price of the base good. The unaware buyer's (subjective) surplus from choosing seller i is

$$x_i \equiv -p_i - (-p^*)$$

Let $D(x_i)$ denote the probability that the buyer chooses seller i.

Firstly, suppose the seller shrouds the add–on. Then the seller chooses p and $\widehat{p}$ to maximize his expected profit

$$(p + (1-\lambda)\widehat{p})\, D(p^* - p)$$

since $E[\widehat{p}] = \overline{p}$ is the equilibrium belief of the aware consumers as we shall see later.

Since, in equilibrium, p^* equals the optimal p, the solution is $p = -(1-\lambda)\overline{p} + \frac{D(0)}{D'(0)}$ and $\widehat{p} = \overline{p}$. Thus the seller gains profit

$$\pi_1 \equiv (p + (1-\lambda)\overline{p})\, D(p^* - p)$$

which confirms the aware consumer's equilibrium belief $E[\widehat{p}] = \overline{p}$.

Secondly, suppose the seller reveals the add-on. Then the fraction of aware buyers becomes $\lambda' > \lambda$.

If the seller chooses $\widehat{p} > e$, then the seller's expected profit is

$$(p + (1-\lambda')\widehat{p})\, D(p^* - p)$$

which is always less than π_1.

If the seller chooses $\widehat{p} \leq e$, then the seller's expected profit is

$$\lambda' (p + \widehat{p})\, D(-p - \widehat{p} + p^* + E[\widehat{p}^*]) + (1-\lambda')\,(p + \widehat{p})\, D(p^* - p)$$

In the solution, $\widehat{p} = e$. Otherwise, the seller can be better off by slightly increasing $\widehat{p}$ and decreasing p by the same increment. In equilibrium, $E[\widehat{p}^*] = e$ as well. Thus the seller's expected profit is

$$\pi_2 \equiv (p + e)\, D(p^* - p)$$

Comparing π_1 and π_2, we see that when $\lambda < 1 - \frac{e}{\overline{p}}$, shrouding is better for the seller. The more unaware buyers in the population, the more likely that the sellers choose to shroud the add-on. However, in the shrouding equilibrium, the unaware buyers are exploited by a high later add-on price $\overline{p}$ which was supposed to be zero.

Furthermore, in the shrouding equilibrium, $p < \frac{D(0)}{D'(0)}$. Thus from the unaware consumer's view, the base good price is "too cheap to be true". So the pricing for the base good in the shrouding outcome is not justifiable.[23]

The inadequacy of market-based solution to the asymmetric information in consumer contracts has been discussed in law literature.[24] An average consumer is not willing to read a standard form contract or λ is low in the model, probably because the contract is too lengthy, the consumer has no clue to understand some contract terms which, say, create a unforeseen legal relationship, or some important information is in the fine print, which is a slip in her mind.

Two other mechanisms may help to deal with it: ex ante intervention by legislatures and ex post reviews by courts. The former mechanism requires a mandatory disclosure

[23] In the Bayesian model version of Gabaix and Laibson (2006), this observation cannot be captured.

[24] See for example Korobkin (2003) and Becher (2008).

rule on the firms. However, this solution fails when the consumers still do not read the contract, or legislative process is inefficiently influenced by interest groups. Moreover, it strongly assumes the legislator's superior knowledge in each industry. The latter mechanism suggests the *unconscionability doctrine*, which refuses to enforce a contract with unconscionable terms. However, this doctrine as applied by common law courts is not defined by status, and thus is too flexible.

4.3 Unaware Agents and Incentive Designs

The classical model of moral hazard between a Principal and an Agent as developed by Mirrlees (1975), Holmström (1979) and Grossman and Hart (1983) assumes that the Agent takes an unobservable action that is typically associated with effort, caution, diligence, time spent, moderation in private consumption, use of efficient technologies, and other decisions. Although this unobservable action is typically complex and multi-dimensional, the basic model integrates all these into the one-dimensional "effort" variable. In reality, however, the Agent typically does not understand all these different dimensions and is unaware of some choice possibilities. For example, in an employment relationship, an employee might be unaware of the possibility of using some set of tools or software to improve her work performance, or another might be unaware of some shirking behavior, such as idling about in "Second Life" in her office, that impairs her performance.

When the Agent is unaware of part of her contracting environment, the standard solution concept for Principal-Agent problems is not satisfactory, since it might not be optimal for the Principal to write all actions of the Agent into the contract and regulate them by means of incentive-compatibility constraints. Incomplete contracts might be a better alternative for the Principal. For instance, if the employer knows that the employee is unaware of some shirking behavior, then it may not be optimal for the employer to regulate this type of activity in the contract, since this makes the employee aware of this type of activity and necessitates the provision of incentives. In the present paper, we therefore consider a generalization of the standard Principal-Agent model in which the Agent is unaware of some dimensions of the effort variable, while being able to optimize over other dimensions.

The paper uses the classical multi-task Principal-Agent model by Holmström and Milgrom (1991) as a starting point. This model is well understood, and captures many important contracting considerations by simple parametrization. We first discuss the optimal incentive contracting problem under the assumption that the Principal knows that the Agent is unaware of some dimension of effort choice and therefore takes an unconscious default action in that dimension that does not respond to incentives. In the model, the Principal designs the incentive scheme and contemplates whether to make the Agent aware of the full problem. We show that if the Agent's default behavior is too lazy or too diligent in this dimension, the Principal will optimally make the Agent aware. On the other hand, if the Agent's default behavior is sufficiently efficient, the Principal will write an incomplete contract where the description of the Agent's action in this dimension is missing. This incompleteness economizes on the costs of incentive provision that would arise if the Principal wanted to regulate the Agent's behavior explicitly. We thus identify a new trade-off in the Principal-Agent problem: the benefit of enlarging the Agent's choice set versus the cost of adding an incentive constraint.

We then extend the analysis to an environment with heterogeneous awareness of Agents, where the Principal cannot distinguish whether the Agent is aware or not. In such an environment, the contract that is optimal for an unaware Agent is not viable, because aware Agents will exploit its low pay-performance sensitivity. Thus the Principal has to screen Agents.

We characterize the solution to the screening problem in terms of two basic parameters of the unawareness problem: the *extent of unawareness* (how many unaware Agents are there?) and the *effect of unawareness* (how does unawareness distort the Agent's action?). Similarly to what happens in the traditional screening problems (see, e.g., Bolton and Dewatripont (2005) and Laffont and Martimort (2002)), unaware Agents are kept to their outside utility, i.e. do not receive any rent from the relationship, and the rent received by aware Agents increases in the extent of unawareness. Hence, the existence of unaware Agents exerts a positive externality on the aware Agents. Differently from standard theory, however, in our model the single-crossing property does not hold. Interestingly, still most of the standard predictions of screening theory hold in this model. The only difference is that efficiency losses can optimally arise for both types of Agents. While unaware Agents always bear too much risk, aware Agents can bear too much or too little risk, depending on the effect of unawareness. Hence, in the parlance of contract theory, there can be "distortion at the top", and this distortion can even go in both directions.

The biggest difference from standard screening problems, however, is that in our problem the extent of unawareness (the population mix) is not exogenous, because the Principal has the option to make the Agent aware of the full problem by proposing a complete contract. In the full analysis of the problem, we study this possibility and show that complete or incomplete contracts can both emerge as optimal. Furthermore, all three different sorts of incomplete contracts obtained earlier can be optimal, separating, pooling, and constrained separating. We further show that the comparative statics of contract incompleteness is surprisingly simple: the larger the extent of unawareness a priori, the more frequent are incomplete contracts at the optimum (where frequency refers to our second key parameter, the measure of the effect of unawareness). This can be interpreted as a self-reinforcing pattern: populations with a large extent of unawareness will operate predominantly with incomplete contracts, thus preserving unawareness. On the other hand, populations with a low extent of unawareness will operate mainly with complete contracts, which eliminate unawareness.

This result is similar to that of Gabaix and Laibson (2006) that firms strategically shroud expensive attributes of their products if and only if there are enough unaware consumers in the market. Gabaix and Laibson (2006) show that such shrouding is stable against competition, while our result shows that shrouding is stable against optimal contracting.

The rest of the paper is organized as follows. The next section reviews the related literature. Section 4.3.2 considers the case of homogeneous unawareness, in which the Principal faces an Agent whom he knows to be unaware of the full contracting problem. Section 4.3.3 introduces heterogeneous unawareness, where the Agent may or may not be unaware of the full problem. Section 4.3.4 discusses the problem of justifiability of contracts from the unaware Agent's point of view. Section 4.3.5 concludes by discussing a number of conceptual points, such as the robustness against competition, communication and so on.

4.3.1 Related Literature

Unawareness and Contract Design

Following Modica and Rustichini (1994), the difficulties of modeling unawareness by conventional economic information theory have been exposed by Dekel, Lipman, and Rustichini (1998) who show that it is impossible to describe non-trivial unawareness in the standard state space model. In response, Heifetz, Meier and Schipper (2006), Li (2008) and Galanis (2007) have proposed theories that circumvent the negative result. The shared feature of these papers is that what is missing in the Agent's mind are not arbitrary points in the state space but a whole dimension. Different from this approach, our work focuses on the Agent's unawareness of her own action set. However, we follow this approach by also assuming that there is a whole dimension of the choice set the Agent is unaware of.

The challenge of incorporating unawareness into dynamic game theory has recently been addressed by Halpern and Rego (2006) and Rego and Halpern (2007) who provide a general setting for studying games with unawareness of actions. The Principal-Agent model that we discuss in our paper uses a simple dynamic game and fits naturally in the approach proposed by these two authors.

Unawareness of actions requires a theory of restricted decision-making by the Agent. Does the Agent optimize over a restricted set? Does she follow some heuristics? As discussed above, our theory must assume that a whole dimension of actions is missing in the Agent's mind. We can therefore address this problem very simply, following Hayek (1968), Vanberg (2002) and others, by assuming that the Agent chooses a default action in the missing dimension and optimizes over the other dimension(s). The default action is an instance of rule-guided or automatic behavior which is not determined by rational choice. There is ample evidence of such behavior in the sociological and psychological literature that documents various forms of automatic versus controlled behavior (see, e.g., Fiske and Taylor (2007)). This rule-guided behavior creates a certain exogenous bias in the Agent's behavior that we take as an important parameter in our comparative statics.

An important conceptual difficulty in understanding unawareness is the interaction between fully aware and unaware contracting parties. Gabaix and Laibson (2006) analyze the interaction between firms and unaware consumers. The consumers who are unaware of later add-on prices are exploited by the firms. In our paper, the Agent is only unaware of her own actions, so there is no issue of exploitation. Filiz-Ozbay (2008) models interaction between a rational insurer and an insuree who is unaware of some contingencies. In contrast, in our paper, what is missing in the Agent's mind is not some future contingencies but her choice possibilities. Closest to our theory in this respect is the work by Eliaz and Spiegler (2006) who study a contract-theoretic model of screening consumers' awareness of their future changed tastes. Eliaz and Spiegler (2008) also study how the firm uses marketing devices to manipulate the consumer's perceived choice set, probably because the consumer is unaware of some products, and analyze the behavioral implications in the context of a competitive market model. In our paper, the Principal is also confronted with Agents of different awareness and designs contracts to exploit these differences. But differently from their work, we focus on the provision of incentives and on how unawareness changes the traditional Principal-Agent paradigm.

Incomplete Contracts

Our work also contributes to the recent literature on the foundations of contract incompleteness. The literature has proposed several reasons why contracting parties may not specify everything that is relevant for the interaction in the contract. Most notable are probably arguments invoking verifiability (Grossman and Hart (1986), Hart and Moore (1990)), signaling (Aghion and Bolton (1987), Spier (1992), Chung and Fortnow (2007)), and explicit writing costs (Dye (1985), Anderlini and Felli (1999), Battigalli and Maggi (2002)). Recent approaches endogenize contractual incompleteness by limited cognition and strategic investment in cognition by the contracting parties (Bolton and Faure-Grimaud (2007), Tirole (2008)). These papers take a less radical approach towards unawareness than Dekel, Lipman and Rustichini (1998), as they assume that agents are aware of the fact that they may be unaware of some relevant elements of the contracting environment. In Gabaix and Laibson (2006) and Filiz-Ozbay (2008), contractual incompleteness arises because better informed agents shroud some contingencies or actions of the informed agents. In the present paper, contracts can be incomplete for the same reason: the Principal strategically shrouds a dimension of action choice by the Agent and only announces a compensation scheme that leaves unaware Agents unaware of the full agency problem. While such shrouding (and hence the distinction between complete and incomplete contracts) is irrelevant in standard Principal-Agent theory, it matters in the context of unawareness, and we characterize its determinants and effects.

Psychological Foundations of Low-Powered Incentives

Next to a number of arguments invoking noisy environments, adverse selection, collusion, reputation, and effort substitution in multi-task environments that show why high-powered incentives can be counter-productive, there is an emerging literature on the psychological foundation of low-powered incentives. Fehr and Schmidt (1999) argue that when agents in teams are inequity averse it is better to lower the incentive power in contracts. Benabou and Tirole (2003) show that high-powered incentives can destroy agents' intrinsic motivation for a task. Benabou and Tirole (2006) further point out that agents may feel shame if they work harder for high-powered incentives. Falk and Kosfeld (2006) document experimental evidence that more restrictions on the agent's action space signal the principal's distrust of the agent, which leads the agent to perform less well. Our work provides a simple framework for the costs of extrinsic motivation in a Principal-Agent relationship: replacing intrinsic effort provision by monetary incentives incurs the cost of second-best contracting, and we show under what conditions on the contracting environment one or the other will be optimal.

4.3.2 The Basic Model

There are two parties, a *Principal* and an *Agent*. The Principal proposes a contract to the Agent to work for him. The Agent's work involves effort in two dimensions, $(t_1, t_2) \in \mathbb{R}^2_+$. In our context, the case of higher-dimensional effort is a straightforward extension. The Agent's effort creates a performance of monetary value $x \equiv t_1 + t_2 + \epsilon$ where ϵ is normally distributed with zero mean and variance σ^2. The effort choices t_1 and t_2 are not observable by the Principal, but x is verifiable. t_1 and t_2 denote different forms of effort spent by the Agent to produce good results that we discuss later in more detail.

By assumption, the Principal remunerates the Agent by a linear compensation rule $w(x) \equiv \alpha x + \beta$. α measures the intensity of the incentives provided to the Agent.

Hence, αx is the incentive pay and β represents the base salary.[25] In standard contract theory, a *contract* is a tuple $(\alpha, \beta, t_1, t_2)$. Although t_1 and t_2 are not observable by the Principal, they can be included in the contract for completeness; their choice, however, must be supported by an appropriate incentive constraint. Alternatively, the parties can write an incomplete contract (α, β) that induces the Agent to choose certain levels of effort by virtue of her incentive constraint. If both parties understand the contracting problem, complete and incomplete contracts are equivalent.

The timing is as follows:

1. The Principal proposes a contract or nothing. If a contract is proposed, the Agent decides whether to accept it. If the Principal proposes nothing or the Agent rejects the contract, then the game is over and each party receives the outside payoff zero.

2. If the Agent accepts the contract, the Agent exerts efforts t_1 and t_2.

3. The outcome of performance x is realized and the contractual compensation is paid.

The Agent's cost of effort can be measured in monetary units and is $C(t_1, t_2) \equiv \frac{1}{2}{t_1}^2 + \frac{1}{2c}{t_2}^2$ with $c > 0$. The smaller c, the more costly the second dimension of effort. The Agent has an exponential von Neuman-Morgenstern utility function over money

$$u(y) \equiv -e^{-y}$$

where we have normalized the coefficient of absolute risk aversion to 1. Since ϵ is normally distributed, we have the standard result that

$$u(CE) = E\left[u(w(x) - C(t_1, t_2))\right]$$

where

$$CE \equiv \alpha(t_1 + t_2) + \beta - C(t_1, t_2) - \frac{1}{2}\sigma^2\alpha^2 \tag{4.1}$$

is the certainty equivalent of the Agent.

The Principal is risk neutral. Hence, his certainty equivalent equals his expected utility

$$E\left[x - w(x)\right] = (1 - \alpha)(t_1 + t_2) - \beta.$$

The *total surplus* of the Principal and the Agent is the sum of their certainty equivalents

$$t_1 + t_2 - C(t_1, t_2) - \frac{1}{2}\sigma^2\alpha^2.$$

The *first best* solution maximizes the total surplus and is given by $t_1^{FB} = 1$, $t_2^{FB} = c$ and $\alpha^{FB} = 0$: the more costly in terms of effort the second task is compared to the first one,

[25] We assume this form of contract because it is simple and captures two important elements of incentive contracting. It is worth noting that Holmström and Milgrom (1987) provide a foundation for this assumption in a dynamic setting.

the less effort is optimally devoted to it. Additionally, the incentive pay is zero, since we have the full insurance result.

The innovation in our paper is the assumption that the Agent is unaware of t_2 before contracting. If she is still unaware of t_2 after contracting, the Agent will choose the *default* effort, or status quo choice $t_2 = \tau \geq 0$ unconsciously in stage 2. The Agent's choice of the default action is not based on rational calculation. τ is only her unconscious rule-guided behavior (see, e.g., Hayek (1967), Vanberg (2002)).

There are two ways of interpreting the notion of unawareness in our context.

The first assumes that the Agent is simply unaware of the possibility of choosing the activities summarized by t_2. This may be the utilization of a certain type of equipment that improves output (in which case the default level τ of not using this equipment is probably inefficiently low) or some form of amenity or perquisite that makes work more pleasant (in which case the default level may be too high or too low, depending on its level).

The second (broader) interpretation of unawareness assumes that the Agent is aware of the activities summarized by t_2, but unaware of their causes and consequences, which Galanis (2007) calls unawareness of theorems. In this case, the Agent chooses τ according to some habits or routines that do not respond to incentives. Examples for this type of activity are unobservable investments into maintaining equipment or the work environment, the Agent's effort in personal customer relations and other forms of personal conduct (where more of these activities correspond to higher τ). Other examples are private emails or phone calls during work or the use of a company car for private ends (where more of these activities correspond to lower τ).[26]

The essential distinction between activities t_1 and t_2 is that the former respond to monetary incentives while the latter do not, unless the Agent is made explicitly aware.

We summarize this explicit communication by the notion of a complete contract $(\alpha, \beta, t_1, t_2)$ in stage 1. If the Agent sees such a contract she will update her awareness and the new dimension of effort choice comes to her mind. On the other hand, if the Principal proposes an incomplete contract where t_2 is missing, the Agent remains unaware of it. Thus, if the Agent is unaware of the full effort problem, complete contracts and incomplete contracts are different instruments.

Under a *complete contract*, the Principal announces (t_1, t_2), and the Agent is aware of t_2. The optimal contract proposed by the Principal is the solution of the following standard

[26]Note that the Agent here is only unaware of the relationship between the costs and benefits of her actions, but knows the actions she is taking. Therefore, the agent knows her restricted utility function as assumed below, although she cannot optimize over all its variables.

problem of multi-task incentive design:

$$\max_{\alpha,\beta,t_1,t_2} (1-\alpha)(t_1+t_2)-\beta$$

$$\text{s.t. } (t_1,t_2)\in \arg\max u(\alpha(t_1+t_2)+\beta-C(t_1,t_2)-\frac{1}{2}\sigma^2\alpha^2) \tag{4.2}$$

$$u(\alpha(t_1+t_2)+\beta-C(t_1,t_2)-\frac{1}{2}\sigma^2\alpha^2)\geq u(0) \tag{4.3}$$

(4.2) is the incentive compatibility constraint for the aware Agent, and (4.3) is her participation constraint.

Because of the simple linear-quadratic form of the problem, the incentive constraint is equivalent to

$$t_1=\alpha, \qquad t_2=c\alpha. \tag{4.4}$$

Since $u(\cdot)$ is strictly increasing, the participation constraint must be binding, and the Principal maximizes the total surplus subjective to the Agent's incentive constraint (4.4). The solution is

$$\alpha^A=\frac{1+c}{1+c+\sigma^2}, \tag{4.5}$$

$$\beta^A=\frac{1}{2}(\sigma^2-1-c)(\frac{1+c}{1+c+\sigma^2})^2, \tag{4.6}$$

$$t_1^A=1-\frac{\sigma^2}{1+c+\sigma^2}, \tag{4.7}$$

$$t_2^A=c-\frac{c\sigma^2}{1+c+\sigma^2}. \tag{4.8}$$

The superscript A means that the Agent is "*A*ware". The Principal's expected profit is

$$\pi^A=\frac{(1+c)^2}{2(1+c+\sigma^2)} \tag{4.9}$$

which is positive. Thus proposing a complete contract is always better than proposing nothing for the Principal.

If the Principal proposes an *incomplete contract*, then the Agent is still unaware of the second effort dimension and will choose the default action τ in stage 2. We assume that the Principal knows τ, interpreted as a typical status quo choice taken by unaware Agents. Thus the Principal solves the following problem:

$$\max_{\alpha,\beta,t_1}(1-\alpha)(t_1+\tau)-\beta$$

$$\text{s.t. } t_1\in \arg\max u(\alpha(t_1+\tau)+\beta-C(t_1,\tau)-\frac{1}{2}\sigma^2\alpha^2) \tag{4.10}$$

$$u(\alpha(t_1+\tau)+\beta-C(t_1,\tau)-\frac{1}{2}\sigma^2\alpha^2)\geq u(0) \tag{4.11}$$

(4.10) and (4.11) are the incentive compatibility constraint and participation constraint, respectively, for the unaware Agent. Since τ is exogenous, there is no incentive-compatibility constraint for t_2.[27] As in the case of awareness, the incentive constraint on t_1 reduces to $t_1 = \alpha$, and the solution to the optimization problem then is

$$\alpha^U = \frac{1}{1+\sigma^2}, \tag{4.12}$$

$$\beta^U = \frac{\tau^2}{2c} - \frac{\tau}{1+\sigma^2} + \frac{\sigma^2 - 1}{2(1+\sigma^2)^2}, \tag{4.13}$$

$$t_1^U = 1 - \frac{\sigma^2}{1+\sigma^2}. \tag{4.14}$$

Observation 4.1 $\alpha^U < \alpha^A$.

If the Principal does not mention t_2 in the contract, then the incentive component plays a less significant role in the wage structure. The intuition is that if the Agent is unaware of one dimension of effort, she is more restricted in adjusting her effort choice. Hence, the outcome x is less sensitive to effort and the Agent's pay should be less sensitive to x.

Observation 4.2 $t_1^U < t_1^A < t_1^{FB}$ *and* $t_2^A < t_2^{FB}$.

Because of the problem of hidden action, the effort levels that the Agent controls are less than the first best levels. Moreover, $t_1^U < t_1^A$, i.e., the aware Agent will work harder even in the dimension in which awareness makes no difference. This is due to Observation 1, $\alpha^U < \alpha^A$: the pay-performance sensitivity plays a more important role in the wage structure when the Agent is fully aware.

When the Agent is unaware, the expected profit of the Principal is

$$\pi^U = \frac{1}{2(1+\sigma^2)} + \tau - \frac{\tau^2}{2c}. \tag{4.15}$$

Combining (4.9) and (4.15), we have

$$\pi^A - \pi^U = \frac{\tau^2}{2c} - \tau + \frac{(1+c)^2}{2(1+c+\sigma^2)} - \frac{1}{2(1+\sigma^2)}. \tag{4.16}$$

The right-hand side of (4.16) is quadratic in τ. Solving this quadratic equation (and ignoring the case of indifference) yields the following proposition.

[27] We assume that the aware Agent and the unaware Agent derive the same utility level from their outside option, say staying at home. In particular, we rule out the possibility that the Agent can improve the value of her outside option when being aware of t_2.

Proposition 4.1 *If the Principal knows that the Agent is unaware of t_2, he optimally proposes $(\alpha^U, \beta^U, t_1^U)$ for values*

$$\tau \in \left(c - \frac{c\sigma^2}{\sqrt{(1+\sigma^2+c)(1+\sigma^2)}}, c + \frac{c\sigma^2}{\sqrt{(1+\sigma^2+c)(1+\sigma^2)}} \right). \tag{4.17}$$

Otherwise, he proposes $(\alpha^A, \beta^A, t_1^A, t_2^A)$.[28]

In other words, the Principal will write an incomplete contract without mentioning t_2 if and only if τ is in the interval given in (4.17).

Proposition 4.1 implies that if the Agent is unaware that she is too lazy or too diligent in some dimension of the effort choice, the Principal will optimally make her aware of this effort dimension. It is quite plausible that if the Agent is unconsciously very lazy, it is better for the Principal to make the Agent aware of it and subject her to explicit incentives. For instance, if the Agent is unaware of using a certain type of equipment to improve the output, the Principal will announce this possibility to the Agent. Interestingly, however, even if the Agent is too diligent, say by cleaning her tools three times after each use, the Principal also has an incentive to make the Agent aware of this dimension of activity. The reason is that the Agent bears the cost even of the actions she undertakes unconsciously. Hence, if τ is too far away from the efficient level, in either case, the Agent's allocation of efforts reduces total surplus, which ultimately hurts the Principal. However, making the Agent aware of this problem also has a cost: it adds an incentive constraint to the Agent's choice problem, with a corresponding reduction of surplus.

Remembering that $t_2^{FB} = c$, Proposition 4.1 reflects the fact that if the Agent's default action is sufficiently close to the first best one, then the Principal will optimally be silent on t_2. If the Principal announces t_2 in the contract, the Principal is forced to provide explicit incentives to the Agent, which in this case is more costly than having the Agent operate at the status-quo level τ. An interesting observation is that even if $\tau \in \left(c - c\sigma^2/\sqrt{(1+\sigma^2+c)(1+\sigma^2)}, t_2^A \right)$, that is, if after making the Agent aware, the Agent will work harder, the Principal still prefers writing an incomplete contract. The reason is that the high effort level t_2^A comes at the expense of a high incentive pay, which hurts the Principal.

As a general rule, *the Principal's decision between a complete and an incomplete contract balances the benefit of enlarging the Agent's choice set against the cost of adding additional incentive constraints.*

The comparative statics of Proposition 4.1 show that when c decreases, the range of τ for which the optimal contract is incomplete shifts to the left and shrinks. Hence, if t_2

[28] Because $\pi^A > 0$, whenever $\pi^U > \pi^A$, we get $\pi^U > 0$. Hence, the Principal always gains from proposing a contract.

becomes more and more costly, it is less probable that the optimal contract is incomplete, and high default effort levels τ are more likely to lead to contract completeness. Similarly, when σ^2 increases, the interval gets larger. Thus the noisier the environment, the more probable it is that the optimal contract is incomplete.

Finally, it should be noted that there is no need for policy intervention to promote the Agent's awareness. Since the Principal maximizes total surplus (subject to the incentive constraint of the Agent), when he prefers an incomplete contract, the total surplus is larger than under a complete contract.

4.3.3 Heterogeneous Awareness

In the previous section, we have assumed that the Principal knows whether the Agent is aware or not. We now generalize the analysis by assuming that the Principal does not know this. Formally, we assume that there are a fraction λ of the Agents who are fully aware (type A) and $1-\lambda$ of the Agents who are unaware of the second dimension of the effort problem (type U), but the Principal cannot distinguish them. To simplify the exposition, we set $c = 1$ (the whole analysis extends to arbitrary $c > 0$). To make the problem interesting, we assume that the default effort level of the unaware Agent lies in the interval (4.17) in which it is in principle better for the Principal to keep the unaware Agent unaware:

$$\tau \in (\tau_{\min}, \tau_{\max}) \equiv \left(1 - \frac{\sigma^2}{\sqrt{(2+\sigma^2)(1+\sigma^2)}}, 1 + \frac{\sigma^2}{\sqrt{(2+\sigma^2)(1+\sigma^2)}}\right) \tag{4.18}$$

Without this assumption, the Agent's unawareness would make her behave so stupidly that the Principal would want to make her aware without any ado. In the previous section, we have identified the contracts $(\alpha^A, \beta^A, t_1^A, t_2^A)$ for the aware Agent and $(\alpha^U, \beta^U, t_1^U)$ for the unaware Agent that the Principal would optimally offer to each of these two types if he knew their type. However, as the following observation shows, if the Principal offers both these contracts Agents of different types do not self-select:

Observation 4.3 *If the Principal proposes the contracts $(\alpha^A, \beta^A, t_1^A, t_2^A)$ and $(\alpha^U, \beta^U, t_1^U)$, the aware Agent will choose $(\alpha^U, \beta^U, t_1^U)$.*

Hence, the aware Agent pretends to be unaware in order to exploit the low pay-performance sensitivity of the U-contract. To show Observation 4.3 note that if the aware Agent chooses $(\alpha^A, \beta^A, t_1^A, t_2^A)$, she receives a certainty equivalent of zero, since her participation constraint (4.3) binds. But if she chooses $(\alpha^U, \beta^U, t_1^U)$, she receives

$$\begin{aligned}
&\max_{t_1,t_2}\{\alpha^U(t_1+t_2)+\beta^U - C(t_1,t_2) - \frac{1}{2}\sigma^2(\alpha^U)^2\} \\
&= \frac{1}{2}(2-\sigma^2)(\alpha^U)^2 + \beta^U \\
&= \frac{(\tau+\sigma^2\tau-1)^2}{2(\sigma^2+1)^2} \geq 0.
\end{aligned}$$

Hence, we need to determine the menu of contracts into which the Agents select themselves according to their type. Yet, there is a second problem. If the Principal proposes a menu of contracts of which one specifies two effort levels (t_1^A, t_2^A), then unaware Agents will become aware of the second dimension t_2, because t_2 is explicitly announced in the menu of contracts. But the Principal can easily circumvent this problem by proposing two incomplete contracts (α^A, β^A) and (α^U, β^U) without mentioning the Agent's effort obligations. As discussed in the previous section, there is no conceptual difference between an incomplete and a complete contract in our setting if the Agent is aware of the full effort problem. The corresponding efforts are automatically implied by the Agent's optimization given the contracts. In fact, from (4.4) with $c = 1$ we know that $t_1^A = t_2^A = \alpha^A$ and $t_1^U = \alpha^U$.

From now on we shall therefore only consider menus $\mathcal{C}^A = (\alpha^A, \beta^A)$, $\mathcal{C}^U = (\alpha^U, \beta^U)$ of incomplete contracts.

The Principal now solves the following fairly standard screening problem, where we already replace the Agent's effort incentive constraint by the corresponding first-order conditions:

$$\max_{\alpha^A,\beta^A,t_1^A,t_2^A,\alpha^U,\beta^U,t_1^U} \lambda[(1-\alpha^A)(t_1^A+t_2^A)-\beta^A] + (1-\lambda)[(1-\alpha^U)(t_1^U+\tau)-\beta^U]$$

$$\text{s.t. } t_1^A = t_2^A = \alpha^A \qquad (ICA')$$

$$t_1^U = \alpha^U \qquad (ICU')$$

$$\alpha^A(t_1^A+t_2^A)+\beta^A - C(t_1^A,t_2^A) - \frac{1}{2}\sigma^2(\alpha^A)^2 \geq 0 \qquad (PCA)$$

$$\alpha^U(t_1^U+\tau)+\beta^U - C(t_1^U,\tau) - \frac{1}{2}\sigma^2(\alpha^U)^2 \geq 0 \qquad (PCU)$$

$$\alpha^A(t_1^A+t_2^A)+\beta^A - C(t_1^A,t_2^A) - \frac{1}{2}\sigma^2(\alpha^A)^2 \geq \max_{t_1,t_2}\{\alpha^U(t_1+t_2)+\beta^U - C(t_1,t_2) - \frac{1}{2}\sigma^2(\alpha^U)^2\} \qquad (ICA)$$

$$\alpha^U(t_1^U+\tau)+\beta^U - C(t_1^U,\tau) - \frac{1}{2}\sigma^2(\alpha^U)^2 \geq \max_{t_1}\{\alpha^A(t_1+\tau)+\beta^A - C(t_1,\tau) - \frac{1}{2}\sigma^2(\alpha^A)^2\} \qquad (ICU)$$

Here, (PCA) and (PCU) are the Agent's participation constraints, for the aware and the unaware type, respectively, and (ICA) and (ICU) the incentive-compatibility constraints that make sure that the aware and the unaware Agent select the appropriate contracts. Although the unaware Agent is rational in the sense that she optimizes her effort choice given the compensation rule and chooses her preferred compensation rule, she does not know why the Principal proposes the menu in question. Thus the assumption of higher order mutual knowledge of the interaction does not hold in our model. We assume that the unaware Agent is not only unaware of the full effort problem, but also boundedly rational in the sense that she cannot infer from the menu that she is unaware of this problem. We come back to this question of the justifiability of contracts in section 4.3.4.

Upon substituting (ICA') and (ICU') into the other expressions of the above problem, the contract design problem becomes

$$\max_{\alpha^A,\beta^A,\alpha^U,\beta^U} \lambda[\left(1-\alpha^A\right)2\alpha^A-\beta^A]+(1-\lambda)[(1-\alpha^U)\left(\alpha^U+\tau\right)-\beta^U] \tag{4.19}$$

$$\text{s.t. } \frac{1}{2}(2-\sigma^2)(\alpha^A)^2+\beta^A\geq 0 \tag{PCA}$$

$$\frac{1}{2}(1-\sigma^2)(\alpha^U)^2+\tau\alpha^U+\beta^U-\frac{\tau^2}{2}\geq 0 \tag{PCU}$$

$$\frac{1}{2}(2-\sigma^2)(\alpha^A)^2+\beta^A\geq\frac{1}{2}(2-\sigma^2)(\alpha^U)^2+\beta^U \tag{ICA}$$

$$\frac{1}{2}(1-\sigma^2)(\alpha^U)^2+\tau\alpha^U+\beta^U\geq\frac{1}{2}(1-\sigma^2)(\alpha^A)^2+\tau\alpha^A+\beta^A \tag{ICU}$$

Figures 4.1 (drawn for $\sigma^2>2$), 4.2 (drawn for $1<\sigma^2<2$), and 4.3 (drawn for $\sigma^2<1$) depict the problem graphically in the space of all contracts (α,β). The figures show the optimal full-information contracts $(\mathcal{C}_F^A,\mathcal{C}_F^U)$ for each type in isolation, as derived in (4.5)-(4.6) and (4.12)-(4.13), respectively, in the last section:

$$\mathcal{C}_F^A=\left(\frac{2}{2+\sigma^2},\frac{2(\sigma^2-2)}{(2+\sigma^2)^2}\right) \tag{4.20}$$

$$\mathcal{C}_F^U=\left(\frac{1}{1+\sigma^2},\frac{\tau^2}{2}-\frac{\tau}{1+\sigma^2}+\frac{\sigma^2-1}{2(1+\sigma^2)^2}\right) \tag{4.21}$$

Here the index F stands for full-information contracts (full information about the Agent's awareness on the side of the Principal). As noted in Observation 4.3, these contracts are not incentive-compatible.

Let

$$u^A(\alpha,\beta)=\frac{1}{2}(2-\sigma^2)\alpha^2+\beta$$

$$u^U(\alpha,\beta)=\frac{1}{2}(1-\sigma^2)\alpha^2+\tau\alpha+\beta-\frac{\tau^2}{2}$$

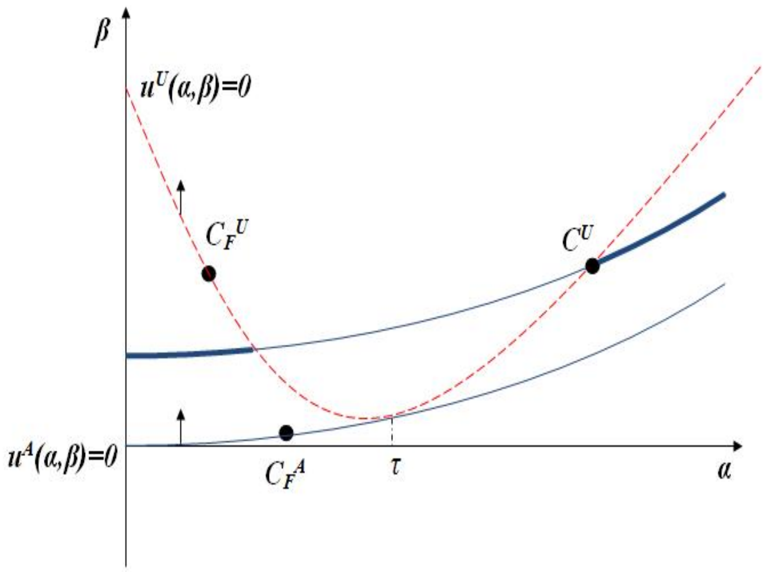


Figure 4.1: Second best solution when $\sigma^2 > 2$

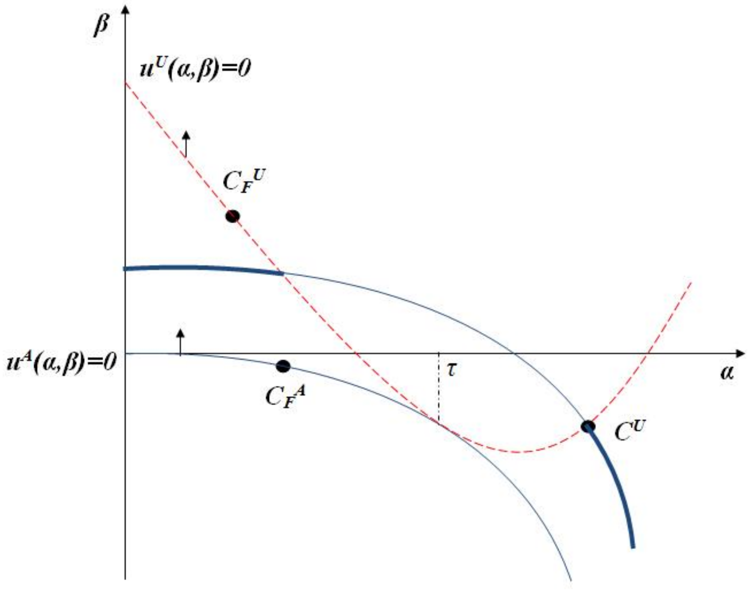


Figure 4.2: Second best solution when $1 < \sigma^2 < 2$

denote the certainty equivalents of the aware and the unaware Agent, respectively, under contracts (α, β). Figures 4.1, 4.2, and 4.3 show the participation boundaries $u^i(\alpha, \beta) = 0$ and an indifference curve $u^A(\alpha, \beta) = const > 0$. Because the indifference curves are quadratic, they can intersect twice. Because of this failure of the single-crossing property to hold, the local incentive analysis of traditional screening problems (see, e.g., Bolton and Dewatripont (2005) and Laffont and Martimort (2002)) will not suffice in this problem.

The figures also show that the participation boundary of the unaware Agent lies above that of the aware Agent, with a point of tangency at $\alpha = \tau$. Starting with this observation, we can simplify the contracting problem by a sequence of arguments that are familiar from standard screening theory.

Lemma 4.1 *(PCA) is redundant in the solution of problem (4.19)-(ICU).*

Proof 4.1 *Direct calculation shows that the right-hand-side of (ICA) satisfies*

$$\frac{1}{2}(2-\sigma^2)(\alpha^U)^2 + \beta^U \geq \frac{1}{2}(1-\sigma^2)(\alpha^U)^2 + \tau\alpha^U + \beta^U - \frac{\tau^2}{2}$$

with equality at $\alpha = \tau$. (ICA) and (PCU) therefore imply (PCA).

Figures 4.1, 4.2 and 4.3 illustrate Lemma 4.1 graphically.

Lemma 4.2 *(ICA) binds at the optimum.*

Proof 4.2 *If the aware Agent's participation constraint (PCA) holds with strict inequality and if (ICA) does not bind, the Principal can raise his profit by slightly lowering β^A without violating any of the constraints. If (PCA) binds, the proof of Lemma 4.1 shows that $\alpha^A = \tau$ and that $(\alpha^U, \beta^U) = (\alpha^A, \beta^A)$. In this case, (ICA) binds trivially.*

In Figures 4.1, 4.2, and 4.3, Lemma 4.2 shows the contract for the aware Agent must lie on the Agent's indifference curve through $\mathcal{C}^U = (\alpha^U, \beta^U)$.

Lemma 4.3 *(PCU) binds at the optimum.*

Proof 4.3 *If (PCA) binds, the proof of Lemma 4.1 shows that (PCU) binds. If (PCA) does not bind, suppose that (PCU) does not bind either. Then the Principal can raise his expected profit by slightly lowering β^A and β^U by the same amount without violating any of the constraints.*

In Figures 4.1, 4.2, and 4.3, Lemma 4.3 shows that $\mathcal{C}^U$ must lie on the participation boundary of the unaware Agent.

Lemma 4.4 *(ICU) is equivalent to $(\alpha^A-\tau)^2 \geq (\alpha^U-\tau)^2$.*

Proof 4.4 *(ICU) is equivalent to*

$$\beta^U-\beta^A \geq \frac{1}{2}(1-\sigma^2)((\alpha^A)^2-(\alpha^U)^2)+\tau(\alpha^A-\alpha^U).$$

By Lemma 4.2, we have

$$\beta^U-\beta^A = \frac{1}{2}(2-\sigma^2)((\alpha^A)^2-(\alpha^U)^2).$$

Hence, (ICU) is equivalent to

$$(\alpha^A)^2-(\alpha^U)^2 \geq 2\tau(\alpha^A-\alpha^U)$$

which is equivalent to

$$(\alpha^A-\tau)^2 \geq (\alpha^U-\tau)^2.$$

Graphically in Figures 4.1, 4.2, and 4.3, Lemma 4.4 says that α^U must be closer to τ than α^A on the horizontal axis.

In Figures 4.1, 4.2, and 4.3, Lemmas 4.1-4.3 imply that the contract $\mathcal{C}^U$ lies on the participation boundary for the unaware Agent, and the contract for the aware Agent must lie on the bold segments of the aware Agent's indifference curve through $\mathcal{C}^U$.

Given the lemmas, problem (4.19)-(ICU) is reduced to the following problem (4.22)-(ICU).

$$\max_{\alpha^A,\beta^A,\alpha^U,\beta^U} \lambda[\left(1-\alpha^A\right)2\alpha^A-\beta^A]+(1-\lambda)[(1-\alpha^U)\left(\alpha^U+\tau\right)-\beta^U] \tag{4.22}$$

$$\text{s.t. } \frac{1}{2}(1-\sigma^2)(\alpha^U)^2+\tau\alpha^U+\beta^U-\frac{\tau^2}{2}=0 \tag{PCU}$$

$$\frac{1}{2}(2-\sigma^2)(\alpha^A)^2+\beta^A=\frac{1}{2}(2-\sigma^2)(\alpha^U)^2+\beta^U \tag{ICA}$$

$$(\alpha^A-\tau)^2 \geq (\alpha^U-\tau)^2 \tag{ICU}$$

Substituting out for β^U and β^A using (PCU) and (ICA) yields a quadratic maximization problem in α^U and α^A subject to the inequality constraint (ICU). In standard screening problems that satisfy the single-crossing property, the incentive constraint (ICU) would not bind, and the (second-best) optimal menu of contracts would generically be separating. Here, things are different in two ways. First, the incentive constraint (ICU) may bind. And second, if it binds this does not necessarily entail pooling ($\alpha^A - \tau = \alpha^U - \tau$), but can also lead to a *constrained separating* outcome ($\alpha^A - \tau = \tau - \alpha^U$). The following proposition shows that all these cases may in fact occur at the optimum.

An important reference point in this proposition is

$$t_{2F}^A = \frac{2}{2+\sigma^2},$$

i.e., the Agent's optimal choice of t_2 under the optimal contract when the Agent is aware of both dimensions of the effort choice. t_{2F}^A has been derived in (4.8) in the last section.[29]

Proposition 4.2 *The solution of problem (4.22)-(ICU) is unique, and there are bounds* $\underline{\tau} \equiv \underline{\tau}(\lambda) < \overline{\tau} \equiv \overline{\tau}(\lambda)$, *with* $\frac{d\underline{\tau}}{d\lambda} > 0$, $\frac{d\overline{\tau}}{d\lambda} < 0$ *for all* $\lambda \in (0,1)$, $\underline{\tau}(0) < \tau_{\min} < \tau_{\max} < \overline{\tau}(0)$ *and* $\underline{\tau}(1) = \overline{\tau}(1) = t_{2F}^A$ *such that:*

1. *If* $\tau < \underline{\tau}$ *or* $\tau > \overline{\tau}$ *the incentive constraint (ICU) is slack and the solution is separating.*
2. *If* $\underline{\tau} \leq \tau \leq t_{2F}^A$ *the incentive constraint (ICU) is binding, and the solution is constrained separating with* $\alpha^A - \tau = \tau - \alpha^U$.
3. *If* $t_{2F}^A \leq \tau \leq \overline{\tau}$ *the solution is pooling.*

The proof is given in the appendix, where we also derive the bounds $\underline{\tau}(\lambda)$ and $\overline{\tau}(\lambda)$ and provide the explicit solutions to the optimal contracts $(\alpha^A, \beta^A, \alpha^U, \beta^U)$.

Figure 4.4 provides a graphical illustration of the different regions identified by Proposition 4.2. The figure focuses on the two key parameters that describe the unawareness problem: τ, the effect of unawareness, and $1-\lambda$, the extent of unawareness. In order to interpret the proposition, we distinguish whether τ is greater or smaller than t_{2F}^A.

If $\tau > t_{2F}^A$ the Agent unconsciously works harder than the optimal full-awareness level. In terms of the graphical description of the contracting problem in $(\alpha, \beta)-$space in Figures 4.1, 4.2, and 4.3, this is the case where the full information contract $\mathcal{C}_F^A$ lies to the left of the point of tangency of the two participation boundaries (which is at $\alpha = \tau$).

Suppose first that λ is sufficiently large, meaning $\lambda > \overline{\tau}^{-1}(\tau)$ in terms of Proposition 4.2 and Figure 4.4. In this case, the intuition can be understood from Figure 4.5, which develops Figure 4.2.[30] The points $\mathcal{C}_F^A$ and $\mathcal{C}_F^U$ are the full-information contracts for the aware and unaware Agents, respectively. The problem for the Principal is to choose a point $\mathcal{C}^U$ on the unaware Agent's participation boundary (PCU) and a point $\mathcal{C}^A$ on the aware Agent's indifference curve through $\mathcal{C}^U$ (ICA) such that α^A is farther from τ than α^U (ICU). Since λ is large, it is optimal for the Principal to have $\mathcal{C}^A$ close to the full-information solution for the aware Agent, and it is no big loss to have $\mathcal{C}^U$ to the right

[29] Note that $2/(2+\sigma^2) \in (\tau_{\min}, \tau_{\max})$ defined in (4.18).
[30] Without loss of generality, our figures focus on the case $1 < \sigma^2 < 2$.

of $\mathcal{C}^A$. This means that (ICU) is slack and there is separation of aware and unaware Agents. The aware Agent's optimal pay-performance sensitivity α^A is not distorted, i.e. equal to the full-information value $\alpha^A_F = 2/(2+\sigma^2)$ identified in the previous section and in (4.20). Yet, the aware Agent receives a larger share of the surplus because her participation constraint does not bind. In Figure 4.5, the loss of the Principal compared to the full-information case is represented by the distance L_A.

On the other hand, the unaware Agent's pay-performance sensitivity α^U is larger than her full-information pay-performance sensitivity, because the aware Agent must be prevented from mimicking the unaware Agent. This efficiency loss is represented by the distance L_U in Figure 4.5. As a result, there is a trade-off of L_A versus L_U for the Principal, that is, *a high base salary for the aware Agent versus an inefficiently high pay-performance sensitivity for the unaware Agent.*

When, on the other hand, λ is small, the intuition is in Figure 4.6. Now there are few aware Agents in the population, so the loss L_A can be larger while L_U should be smaller. This means that $\mathcal{C}^U$ is closer to $\mathcal{C}^U_F$ and $\mathcal{C}^A$ lies on a higher indifference curve of the aware Agent. But the unaware Agent's incentive constraint (ICU) imposes a limit on how much $\mathcal{C}^U$ can be moved towards $\mathcal{C}^U_F$: $\mathcal{C}^U$ must lie (weakly) to the right of $\mathcal{C}^A$. The optimum therefore pools the aware and unaware Agents and furthermore distorts the aware Agent's pay-performance sensitivity inefficiently below α^A_F. Compared to the trade-off of L_A versus L_U in Figure 4.5, now, when λ is small, there is a new efficiency loss: insufficient risk bearing by the aware Agent because of a reduced pay-performance sensitivity, which is represented by the distance L'_A in Figure 4.6.

The second parameter constellation to consider is the case $\tau < t^A_{2F}$, which means that the Agent unconsciously works less hard than the optimal full-awareness level. In contrast to the graphical description of the contracting problem in $(\alpha, \beta)-$space in Figures 4.1, 4.2, and 4.3, this is the case where the full information contract $\mathcal{C}^A_F$ lies to the right of the point of tangency of the two participation boundaries (which is at $\alpha = \tau$).

When λ is large, the intuition is as in the case $\tau > t^A_{2F}$ described above and given in Figure 4.7. Since there are many aware Agents, the contract $\mathcal{C}^A$ should be close to the full-information contract $\mathcal{C}^A_F$, which does not conflict with the requirement that α^U must be closer to τ than α^A (the (ICU) constraint). The Principal's loss L_U on the unaware Agent is relatively large, and the solution is separating as before. The pay-performance sensitivity α^A for the aware Agent is equal to the full-information level and her base salary β^A is relatively low, but the pay-performance sensitivity α^U for the unaware Agent is inefficiently large.

When λ is small, the reasoning changes and the intuition is given in Figure 4.8. Now the Principal wants to keep the loss L_U small and can tolerate a higher loss L_A. Moving $\mathcal{C}^U$ towards $\mathcal{C}^U_F$ reduces the unaware Agent's incentive pay α^U towards the full-information level (without reaching it) and increases the aware Agent's base salary β^A. This movement along the unaware Agent's participation boundary is again restricted by the incentive constraint (ICU), but instead of pooling this implies a "far" separating solution. The reason is that in order to keep the unaware Agent from mimicking the aware one, the aware Agent's pay-performance sensitivity α^A must be sufficiently large compared to α^U (because $\mathcal{C}_F^U$ and $\mathcal{C}_F^A$ lie on different sides of τ). Since the loss from inefficiently distorting the pay-performance sensitivity of the aware Agent (L'_A in the figure) matters less in the aggregate because λ is small, it is optimal to do so, necessitating an increase of α^A for an even stronger separation of the two types. Now the unaware Agent's pay-performance sensitivity α^U is close to the full-information level, while the aware Agent's pay-performance sensitivity α^A is distorted upwards compensated by a high base salary.

Hence, when the unaware Agent works hard (τ large), it can be optimal to distort the aware Agent's incentive pay downward (in order to pool both types of Agent), and if the Agent unconsciously provides relatively little effort (τ small), it can be optimal to distort the aware Agents' incentive pay upward (in order to separate them).

Interestingly, except for this latter finding, despite of the failure of the single-crossing property to hold, most of the preceding analysis is as in standard screening models. For example, the "bad" type (the aware Agent) gets a positive rent, while the "good" type is kept to her reservation utility. Also, here as in standard models, the incentive constraint of the "bad" type binds. Furthermore, the aware Agents' rents are decreasing in their population share λ. On the other hand, the failure of the single-crossing property to hold implies that efficiency losses (i.e. distortions in α) can arise for both types, and the incentive-compatibility constraints of both types can bind in a separating solution.

Yet, the analysis in this section so far is incomplete. In contrast to the standard screening problem, the Principal has another option: making the unaware Agent aware. Thus λ cannot be regarded as an exogenous variable. To wit, in the full-information solution $(\mathcal{C}_F^A, \mathcal{C}_F^U)$ derived in Section 4.3.2 it is not wise for the Principal to make the unaware Agent aware when condition (4.18) holds. But if the Principal does not know whether the Agent is aware, it might be better for the Principal to announce the full effort problem, in order to avoid the screening costs associated with the allocation in Proposition 4.2.

As noted before, the aware Agent's rents are decreasing in λ. Yet, these rents are paid with probability λ, hence, it is not clear what the impact of λ on total rents is, and whether this affects the Principal's announcement decision. The following proposition provides a surprisingly clear-cut answer.

Proposition 4.3 *There are bounds $\tau_L \equiv \tau_L(\lambda) < \tau_R \equiv \tau_R(\lambda)$, with $\frac{d\tau_L}{d\lambda} > 0$, $\frac{d\tau_R}{d\lambda} < 0$, $\tau_L(1) = \tau_R(1) = t_{2F}^A$, $\tau_L(0) = \tau_{\min}$ and $\tau_R(0) = \tau_{\max}$, such that the Principal optimally proposes the contracts identified in Proposition 4.2 if $\tau \in (\tau_L, \tau_R)$, and makes the Agent* *aware otherwise.*

The proof of the proposition is in the appendix, where we also derive the bounds τ_L, τ_R explicitly. Figure 4.9 illustrates Proposition 4.3 in terms of the τ-λ diagram of Figure 4.4. In the shaded area, the solution of problem (4.22)-(ICU) is dominated by making all Agents aware.

In the extreme case $\lambda = 0$, we are back to the case of the unaware Agent of Section 2. In terms of Figure 4.6 and 4.8, the loss L_U can be reduced to zero at no cost, and indeed the condition $\tau \in (\tau_L(0), \tau_R(0))$ of the proposition is the same as the no-announcement condition (4.18).

In the other extreme case $\lambda = 1$, the Principal always prefers making the Agent aware, since this means offering the optimal contract for the aware Agent and there are only aware Agents in the population. In the knife-edge case $\tau = t^A_{2F}$, both aware Agent and unaware Agent choose the same level of effort in the second dimension. Thus $\tau = t^A_{2F}$ is the only situation where the Principal is indifferent between making them aware and screening them when $\lambda = 1$.

In the non-degenerate case $0 < \lambda < 1$, the Principal trades off two types of contracting costs against the alternative of offering the full-awareness contract $\mathcal{C}^A_F$ to all Agents. Remember from Section 3 that for $\tau \in (\tau_{\min}, \tau_{\max})$ the Principal prefers to keep unaware Agents unaware if this can be done at no cost. The reason is that the Principal this way can economize on incentive pay, which is costly because of hidden actions. However, since the Principal cannot tell the Agents apart, there are further costs of contracting with them. In one type of contract (standard separating contracts), the Principal gives aware Agents the efficient contract in terms of risk-bearing, but provides excessive base pay and distorts the risk-profile of the unaware Agents. In the other type of contract (pooling or constrained separating contracts), unaware Agents are allowed to bear less risk, but aware Agents are treated inefficiently. The former type of distortion is optimal if there are many aware Agents (λ large), the latter if there are many unaware Agents (λ small). Since under both types of contracts, the Principal profits more from the unaware Agents than from the aware ones, the larger the population of aware Agents, the more likely it is that the Principal prefers making all Agents aware.

The resulting contracts have a self-reinforcing pattern. If there are many unaware Agents in the population, optimal contracts will tend to shroud the work dimension in question. If many Agents are aware of both effort dimensions, optimal contracts tend to make everybody aware, thereby eliminating unawareness altogether. The dispersion in contracting outcomes depends on the default level τ that Agents deploy when unaware.

This self-reinforcing pattern is similar to the finding in Gabaix-Laibson (2006) where the shrouding of product attributes is shown to be optimal if there are many unaware consumers in the market. In their model of competitive markets, this happens because educating unaware consumers allows them to profit from the firms' otherwise competitive pricing, thus exposing firms to losses if there are many unaware consumers. In our model the reason is very different: shrouding is optimal because it economizes on the costs of providing effort incentives. The resulting contract features lower-powered incentives than the usual full-awareness contract. If the degree of awareness in the population (λ) increases, it will at one point become optimal to make all Agents aware and use high-powered incentives.

4.3.4 Justifiability of Contracts

Filiz-Ozbay (2008), Ozbay (2008), and Heifetz, Meier and Schipper (2009) argue that in the context of unawareness a reasonable equilibrium concept should include the requirement that the Agent thinks that the contract is justifiable, in the sense that the contract is optimal for the Principal also from the Agent's point of view. In this section, we explore the ramifications of this added requirement for our analysis.

First, it is simple to see the solution of the basic contracting problem in section 4.3.2 is justifiable in this sense. If the optimal contract is complete the Agent is aware, and the problem reduces to the standard Principal-Agent problem, the solution of which is even robust to common knowledge of rationality and the contractual setting. If the optimal contract is incomplete, the Agent remains unaware and unconsciously chooses $t_2 = \tau$. Then the Agent's objective function (4.1) includes a fixed-cost element, and again the contract is optimal in the Agent's mind.

However, the solution of the heterogeneous awareness problem in section 4.3.3 is not necessarily justifiable. When the Principal prefers incomplete contracts (the case $\tau \in (\tau_L, \tau_R)$ in Proposition 4.3), the unaware Agent does not understand why there are two different contracts (separating or constrained separating) or a single contract designed the way it is (pooling solution). If we require the contract to be justifiable, the solution is either the complete contract outcome (full awareness outcome) or the incomplete contract $\mathcal{C}_F^U$ that makes sense for the unaware Agent.[31] The additional justifiability constraint therefore reduces the Principal's profit from proposing incomplete contracts and makes him more likely to make all Agents aware. This is because we add an additional justifiability constraint in the problem (4.22)-(ICU). Furthermore, with $\mathcal{C}_F^U$, the aware Agent obtains a maximal rent as noted in Observation 4.3.

Yet, it can be argued that justifiability is too strong a requirement. If one acknowledges that the model necessarily only describes a simplified snapshot of a full (highly multi-dimensional) contracting problem, it may well be reasonable to assume that the Agent

[31] Here, we assume that the aware Agent understands the full contracting problem as well as the Principal. If the aware Agent does not know that other Agents may be unaware, this contract is not justifiable either.

does not want or need to understand the reason for what she sees, as long as what she chooses is optimal for her.

In what follows we therefore propose a weaker justifiability restriction than that of Filiz-Ozbay (2008) and Ozbay (2008). In the spirit of Bolton and Faure-Grimaud (2007) and Tirole (2008), we assume that if the observed contracts are not justifiable, the unaware Agent becomes aware that something might be wrong with her view and starts thinking about it.[32] This cognitive effort leads to her full awareness with probability δ. With probability $1-\delta$ the Agent remains unaware and chooses (one of) the proposed contracts without further ado. This extension can be easily integrated into the analysis of the preceding section. In fact, after seeing a non-justifiable contract, i.e., a menu of incomplete contracts that is different from the single contract $\mathcal{C}_F^U$, the fraction of aware Agents increases to $\lambda' \equiv \lambda + (1-\lambda)\delta$. Hence, non-justifiable contracts promote awareness. Thus making all Agents aware is more likely to be better than non-justifiable contracts according to Figure 4.9, as both τ_L and τ_R are monotone.

Now there are three alternatives for the Principal: (i) making all Agents aware, (ii) proposing $\mathcal{C}_F^U$ alone, (iii) proposing the menu of contracts identified in Proposition 4.2, with the fraction of aware Agents replaced by $\lambda'(>\lambda)$.

From Figure 4.9, we see that now the incomplete contracting solution (iii) becomes less attractive than the complete contracting solution (i), as λ' becomes larger. Yet, alternatives (ii) and (iii) may be optimal in some circumstances. For example, when λ is small, τ is far from t_{2F}^A and δ is large, proposing $\mathcal{C}_F^U$ alone (alternative (ii)) is optimal. In Figure 4.9, we can see that in this case non-justifiable contracts (alternative (iii)) tend to be worse than the complete contracts (i) since λ' is large. Furthermore, proposing $\mathcal{C}_F^U$ alone is better than alternative (i): since λ is small, the loss from the rent of the aware Agent is small for the Principal. On the other hand, when δ is small enough, alternative (iii) can be optimal, as λ' is not significantly greater than λ.

4.3.5 Concluding Remarks

This paper provides a model of incentive design for unaware Agents. It is worth pointing out the following issues.

Competition. The result is not significantly changed if we introduce Bertrand-competition among homogeneous Principals. In this case, the Principal proposes the same contracts as before, except that the base salaries β^A and β^U are both increased by an amount such that the Principal earns zero expected profits. Since this leads to the maximal profit for each Principal, no one has an incentive to deviate from it. In particular, if all Principals propose incomplete contracts, no one has an incentive to make all Agents aware, since

[32] The strategic interaction is based on the solution concepts by Chen and Zhao (2009).

this only decreases their profits. Thus competition cannot promote awareness of the Agents.

Communication-Proofness. We have shown that under certain conditions, the Principal prefers to leave the Agent unaware of the full contracting problem. Interestingly, even the aware Agent has no strict incentive to make her unaware colleague aware through communication. First, the aware Agent will not do so before contracting, because, in the solution, the unaware Agent exerts a positive externality on the aware Agent by conferring a positive rent on her. Second, also after contracting, the aware Agent has no incentive to do so either, because making the unaware Agent aware cannot create any extra rent for herself but only hurts the Principal. Hence, the analysis in section 4.3.3 is robust to the possibility of internal communication among Agents.

Dynamic stability. Our analysis has been static, but it lends itself to an interesting dynamic interpretation. In Proposition 4.3 we have shown that the optimal contracts are more likely to leave unaware Agents unaware for smaller λ (where the measure of Agents is taken with respect to τ). Hence, the more Agents are likely to be unaware, the more will remain unaware after contracting, and vice versa. This suggests a certain stability of unawareness. This observation may be particular important as it suggests that deviations from full rationality are not necessarily doomed to die out in the long run.

Welfare Implication of Public Announcements. Is there room for a benevolent policy maker to improve the outcome through promoting the Agent's awareness?[33] In section 4.3.2 we have argued that when there are only unaware Agents in the population, a public policy of making Agents aware is not welfare-enhancing. The reason is that such an announcement would force the Principal to provide explicit incentives to the Agent, which is costly. Since the Principal maximizes total surplus, the fact that he does not choose to make the Agent aware shows that the costs outweigh the benefits. In section 4.3.3, when there are heterogeneous Agents, this conclusion still holds. If the Principal prefers to leave the Agent unaware, the outcome Pareto-dominates the outcome of making all Agents aware. To see this point, note that when the Principal prefers to leave the Agent unaware, he must earn a higher expected profit than making her aware. Furthermore, the aware Agent earns a positive rent while she would earn zero rents if all Agents were made aware. Finally, the unaware Agent earns a zero rent in any case. Thus there is no need for the policy maker to intervene in the heterogeneous environment as well. Of course, this no-intervention recommendation crucially depends on the assumptions that the Principal knows all the choice possibilities of the Agent.

Slip-of-Mind or Clueless Unawareness. Board and Chung (2008) distinguish "slip-of-

[33]This has been suggested, e.g., by Korobkin (2003) in the context of unawareness about other agents' actions.

mind unawareness" and "clueless unawareness". Under the former, the Agent becomes aware of the full problem as soon as she sees its description in the contract, under the latter she is "hard-wired" in her choice of the default action. Our model can encompass both forms of unawareness if we model the cost of communication a complete contract. Then the two cases of Board and Chung (2008) represent two extremes, one with zero costs, and the other with infinitely high cost. Compared to our model, the existence of the cost only makes the alternative of making all Agents aware less attractive for the Principal.

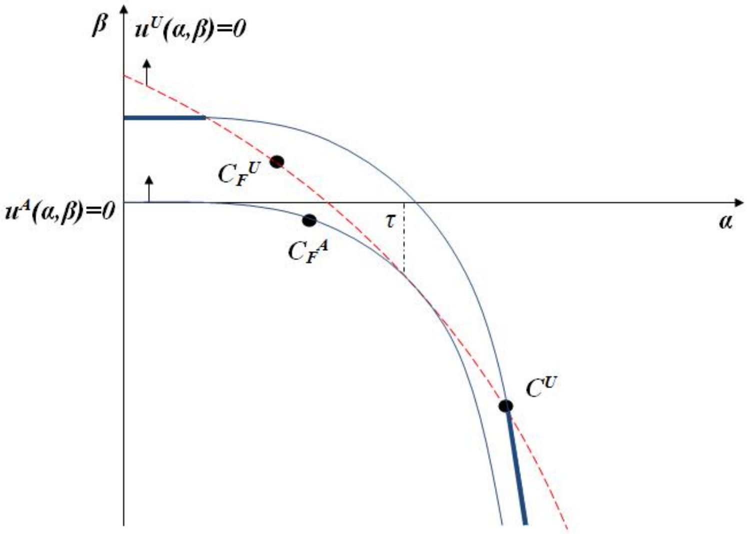


Figure 4.3: Second best solution when $\sigma^2 < 1$

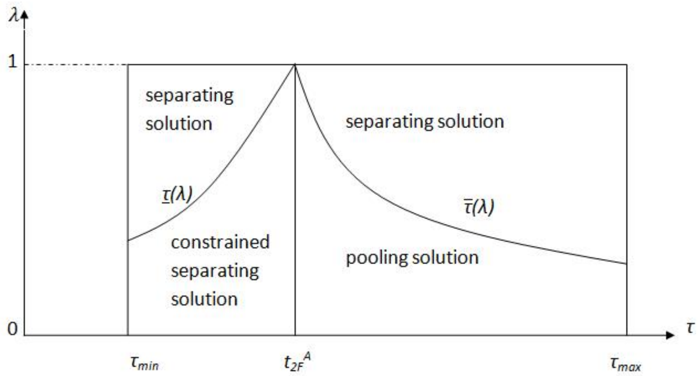


Figure 4.4: Graphical illustration of the three regimes of Proposition 4.2

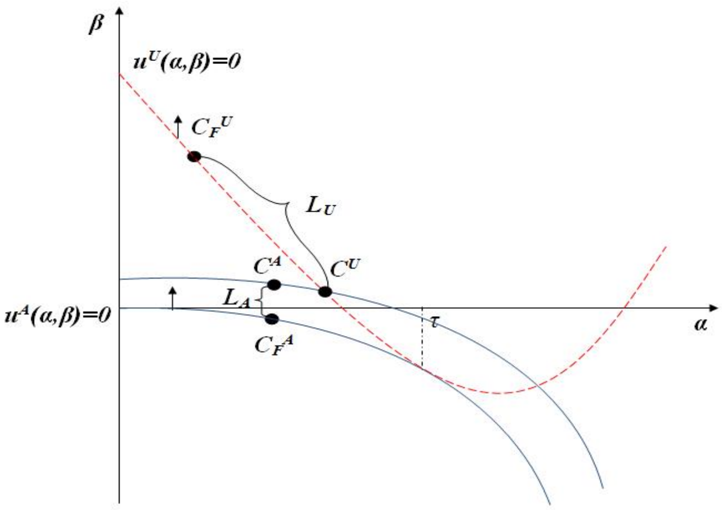


Figure 4.5: Solution of problem (4.22) when λ is large and $\tau > t^A_{2F}$

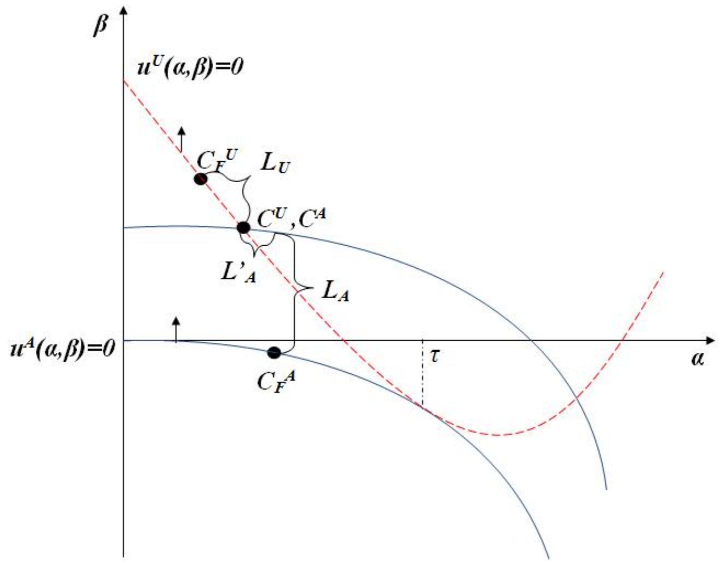


Figure 4.6: Solution of problem (4.22) when λ is small and $\tau > t^A_{2F}$

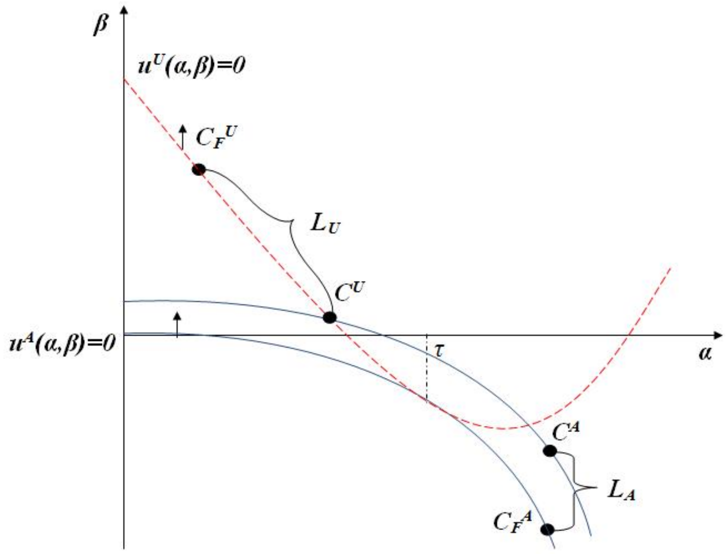


Figure 4.7: Solution of problem (4.22) when λ is large and $\tau < t^A_{2F}$

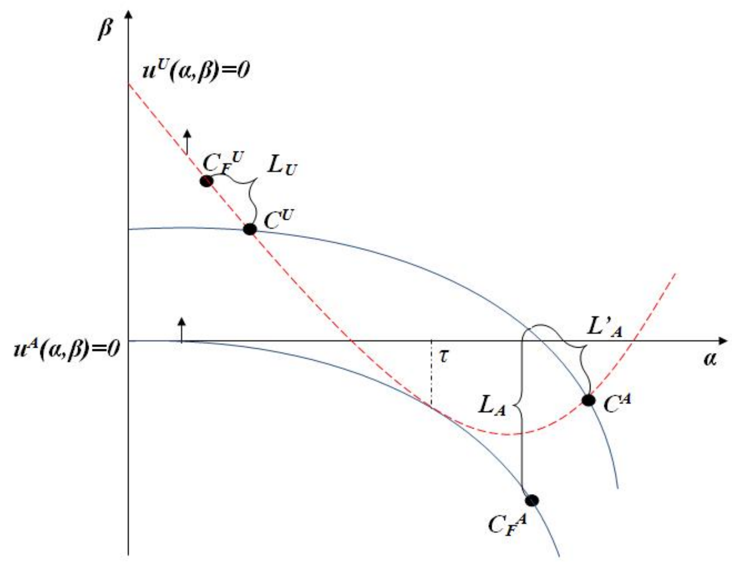


Figure 4.8: Solution of problem (4.22) when λ is small and $\tau < t_{2F}^{A}$

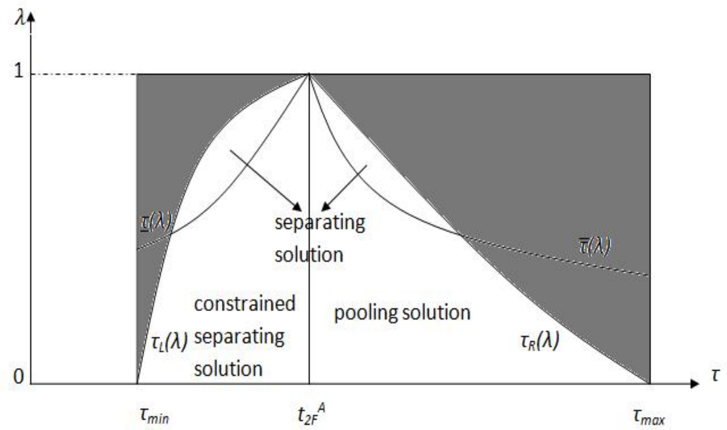


Figure 4.9: Graphic illustration of proposition 4.3

Chapter 5

Conclusion

Besides unforeseen contingencies and unawareness of actions, in games, the agent may not perceive perfectly his utility as well. Sometimes, agents' preferences are time inconsistent. Further related works are self-awareness of changing preferences and intra-personal contracts.

Economics is based on the research paradigm of methodological individualism (Buchanan, 1979), that is, our understanding and explanation should be in terms of the behaviors of individuals. However, what is an individual? Are you tomorrow still you today, especially when your preference is changing over time? For most individuals, their preferences are dynamically inconsistent or time inconsistent which means they do not always make the same choice between two alternatives when they are asked at different times. The multi-selves approach which originates from Stroz (1955) and Laibson (1997) means, by splitting one's personality, we regard an individual as a collection of many selves. Each self has his own preference. Then the left work for us is to solve an intra-personal game of interaction among different selves (players).

Thus self-awareness and intra-personal contracts would be a promising research area in the future.

Appendix A

Appendix

A.1 Appendix

A.1.1 Proof of Proposition 3.1

In case 1, we have that

$$\begin{aligned}&TS_A - TS_U\\&= x\left(\int_{\alpha-1}^{\widehat{a}}\left(g\left(\widetilde{a}_N, a\right) - g\left(\alpha, a\right)\right)da + \int_{\widehat{a}}^{\alpha}\left(g\left(a, a\right) - g\left(\alpha, a\right)\right)da\right)\\&-\left(\left(1-x\right)+\left(\alpha-\widehat{a}\right)\right)c.\end{aligned}$$

$TS_A - TS_U$ is negative if $\widetilde{a}$ is small enough. The other sufficient condition for $TS_A - TS_U < 0$ is $3\widehat{a} + \alpha - 3 > 0$, which holds for sufficiently large α.

Conversely, $TS_A - TS_U$ is positive for sufficiently small α. To show this possibility, consider the extreme case of condition (3.4) in case 1, i.e., $\alpha = 1 - 2\sqrt{c}$. We have that $\widehat{a} = 1 - \alpha$, and

$$TS_A - TS_U \mid_{\alpha=1-2\sqrt{c}} = x\left(1-\alpha\right)\frac{2\alpha^2 - \left(2\alpha-1\right)^2}{8} - \left(\left(1-x\right)+\left(2\alpha-1\right)\right)c,$$

which is more likely to be positive for larger x. Hence, let $x = 1$. We have that

$$TS_A - TS_U \mid_{\alpha=1-2\sqrt{c},x=1} = \left(1-\alpha\right)\frac{2\alpha^2 - \left(2\alpha-1\right)^2}{8} - \left(2\alpha-1\right)\frac{\left(1-\alpha\right)^2}{4}.$$

Since Assumption 3.1 becomes

$$\frac{\left(1-\alpha\right)^2}{4} < \frac{4\alpha-1}{16}$$

which is nothing but $\frac{1}{2} < \alpha < \frac{5}{2}$, and $\alpha = 1 - 2\sqrt{c}$, we have that $\frac{1}{2} < \alpha < 1$. Direct calculation gives

$$TS_A - TS_U \mid_{\alpha=\frac{1}{2},x=1} = \frac{1}{128} > 0.$$

Since $TS_A - TS_U$ is continuous function of x, α and c, we know that $TS_A - TS_U$ is positive for (x, α, c) close enough to the point $(1, 1/2, 1/16)$.

Regarding the consumer surplus, we have that

$$CS_A - CS_U = x\left(\int_{\alpha-1}^{\widehat{a}} f\left(\widetilde{a}_N, a\right) - f\left(\alpha, a\right) da + \int_{\widehat{a}}^{\alpha} f\left(a, a\right) - f\left(\alpha, a\right) da\right).$$

We have $f(a,a) - f(\alpha, a) > 0$ since $\alpha > a$. Therefor, we have $\int_{\widehat{a}}^{\alpha} f(a,a) - f(\alpha,a)\, da > 0$. Moreover,

$$\begin{aligned}
& sign\left(\int_{\alpha-1}^{\widehat{a}} f\left(\widetilde{a}_N, a\right) - f\left(\alpha, a\right) da\right) \\
&= sign\left(\frac{\left(\widehat{a}^2 - (\alpha-1)^2\right)\left(\widetilde{a}_N - \alpha\right)}{4} - \frac{3\left(\widehat{a} - (\alpha-1)\right)\left(\widetilde{a}_N^2 - \alpha^2\right)}{8}\right) \\
&= sign\left(-\frac{\left(\widehat{a} + (\alpha-1)\right)}{4} + \frac{3\left(\widetilde{a}_N + \alpha\right)}{8}\right) \\
&= sign\left(-2\left(\widehat{a} + (\alpha-1)\right) + 3\left(\frac{\widehat{a} + \alpha - 1}{2} + \alpha\right)\right) \\
&= sign\left(5\alpha + 1 - \widehat{a}\right) > 0.
\end{aligned}$$

Thus $CS_A - CS_U$ is always positive.

In case 2, it is straightforward to show that

$$\begin{aligned}
& \frac{\partial\left(TS_A - TS_U\right)}{\partial x} \\
&= \frac{1}{8}\left(\left(2 + \alpha - 4\sqrt{c} - \left(\alpha^2 - 2\alpha\sqrt{c} + 4c\right)\right)2\sqrt{c} + \left(2 + \alpha - 4\sqrt{c}\right)(1-\alpha)\alpha\right) \\
&+ \left(1 - \left(\alpha - 2\sqrt{c}\right)\right)c \\
&> \frac{1}{8}\left(2 + \alpha - 4\sqrt{c} - \left(\alpha^2 - 2\alpha\sqrt{c} + 4c\right)\right)2\sqrt{c}
\end{aligned}$$

(since $\alpha > 2\sqrt{c}$), which is greater than

$$\frac{1}{8}\left(2 + \alpha(1-\alpha) - 4\sqrt{c}\right)2\sqrt{c} > 0$$

since $c < \frac{1}{16}$, because in case 2 we have $2\sqrt{c} < \alpha < 1 - 2\sqrt{c}$.

Lastly, we have

$$CS_A - CS_U = x\int_{\widehat{a}}^{\alpha}\left(f(a,a) - f(\alpha,a)\right) da > 0.$$

A.1.2 Proof of Lemma 3.2

We define another $\sigma-$algebra $\mathcal{B}(K_0)$ that is the smallest $\sigma-$algebra over Ω such that $X \setminus K_0 \in \mathcal{B}(K_0)$, $X^C \setminus K_0 \in \mathcal{B}(K_0)$ and, for all $\omega \in K_0$, $\{\omega\} \in \mathcal{B}(K_0)$. Since the collection of $X \setminus K_0$, $X^C \setminus K_0$ and $\{\omega\}$ for all $\omega \in K_0$ is a partition of Ω, it is the finest partition of Ω with respect to $\mathcal{B}(K_0)$. It is left to show that $\mathcal{L}(K_0) = \mathcal{B}(K_0)$. It is equivalent to show that

1. $X \setminus K_0 \in \mathcal{L}(K_0)$,

2. $X^C \setminus K_0 \in \mathcal{L}(K_0)$ and

3. $X \in \mathcal{B}(K_0)$.

Firstly, since, for all $\omega \in K_0$, $\{\omega\} \in \mathcal{L}(K_0)$, we have $K_0 \in \mathcal{L}(K_0)$. Thus $K_0^C \in \mathcal{L}(K_0)$. Moreover, $X \in \mathcal{L}(K_0)$ implies $X \cap K_0^C \in \mathcal{L}(K_0)$. That is, $X \setminus K_0 \in \mathcal{L}(K_0)$.

Secondly, by the same argument above, we can show $X^C \setminus K_0 \in \mathcal{L}(K_0)$.

Finally, since $\{\omega\} \in \mathcal{B}(K_0)$ for all $\omega \in K_0$, we have $X \cap K_0 \in \mathcal{B}(K_0)$. Moreover, $X \setminus K_0 \in \mathcal{B}(K_0)$. Thus $X = (X \cap K_0) \cup (X \setminus K_0) \in \mathcal{B}(K_0)$.
□

A.1.3 Proof of Proposition 3.3

Firstly, we show the "only if" part. Suppose C^K is not vague but exploitative. Then $\mu^K(\{\omega\}) = \mu(\{\omega\})$ for all $\omega \in \Omega$ by Assumption 3.2. Since C^K is the optimal solution, the participation constraint of the problem (3.11) is satisfied. C^K is therefore not exploitative, a contradiction.

Secondly, we show the "if" part. Suppose C^K is vague but not exploitative. We have

$$\sum_{\omega \in \Omega} \mu(\{\omega\}) u_A(\omega, C^K(\omega)) \geq \sum_{\omega \in \Omega} \mu(\{\omega\}) u_A(\omega, \overline{s}). \tag{A.1}$$

We now define another contract $C^{K'}$ such that $C^{K'}$ is not vague ($K' = \Omega$) and $C^{K'}(\omega) = C^K(\omega)$ for all $\omega \in \Omega$. (By Assumption 3.3, such $C^{K'}$ exists.) Thus $C^{K'}$ gives the principal the same objective expected utility level as C^K does. Moreover, since $C^{K'}$ is not vague, we have $\mu = \mu^{K'}$ by Assumption 3.2. Thus the participation constraint of the problem (3.11) is satisfied because this constraint is nothing but (A.1). Hence $C^{K'}$ is also an optimal contract. However, $C^{K'}$ is less vague than C^K. It contradicts with the tie-breaking rule in Assumption 3.4.
□

A.1.4 Proof of Lemma 3.3

Let $\alpha = \underline{\alpha}$. By equation (3.20), we have $u(w_0 - t_a) = u(w_1 - t_c)$. Combining (3.18), we obtain $(p+q)u(w_0 - t_a) - pu(w_0) - qu(w_1) = 0$. $u(w_1) > u(w_0)$ yields $u(w_0 - t_a) > u(w_0)$. By (3.19), we therefore have $\underline{\alpha} < 1 - p - q$.
□

A.1.5 Proof of Lemma 3.4

We denote the solution of problem (3.14) as a function of α: $t_a(\alpha)$, $t_b(\alpha)$ and $t_c(\alpha)$. By equation (3.17), we have $u(w_0 - t_a(\alpha)) = u(w_1 - t_c(\alpha))$. Combining (3.15), we obtain

$$(p+q)u(w_0 - t_a(\alpha)) + \alpha[u(w_0 - t_b(\alpha)) - u(w_0)] = pu(w_0) + qu(w_1). \tag{A.2}$$

If $\alpha = \underline{\alpha}$, we then have

$$(p+q)u(w_0 - t_a(\underline{\alpha})) = pu(w_0) + qu(w_1). \tag{A.3}$$

Combining (A.2) and (A.3), we get

$$(p+q)[u(w_0 - t_a(\alpha)) - u(w_0 - t_a(\underline{\alpha}))] + \alpha[u(w_0 - t_b(\alpha)) - u(w_0)] = 0. \tag{A.4}$$

Firstly, let $\alpha > \underline{\alpha}$. Then, by equation (3.16) and (3.19), we get $\frac{u'(w_0-t_a(\alpha))}{u'(w_0-t_b(\alpha))} > \frac{u'(w_0-t_a(\underline{\alpha}))}{u'(w_0)}$. It implies $\frac{u'(w_0-t_a(\alpha))}{u'(w_0-t_a(\underline{\alpha}))} > \frac{u'(w_0-t_b(\alpha))}{u'(w_0)}$. Suppose $u(w_0 - t_b(\alpha)) \leq u(w_0)$. Then $u'(w_0 - t_b(\alpha)) \geq u'(w_0)$. Thus $\frac{u'(w_0-t_a(\alpha))}{u'(w_0-t_a(\underline{\alpha}))} > 1$. It implies $u(w_0 - t_a(\alpha)) < u(w_0 - t_a(\underline{\alpha}))$. But it makes the proved equation (A.4) impossible. Thus we must have $u(w_0 - t_b(\alpha)) > u(w_0)$.

Secondly, we can show $u(w_0 - t_b(\alpha)) < u(w_0)$ if $\alpha < \underline{\alpha}$ by the same argument. Thus $\alpha \geq \underline{\alpha}$ if and only if $t_b \leq 0$.
□

A.1.6 Proof of Lemma 3.5

Let $\begin{pmatrix} f(t_a,t_b,t_c,\alpha) \\ g(t_a,t_b,t_c,\alpha) \\ h(t_a,t_b,t_c,\alpha) \end{pmatrix}$
$\equiv \begin{pmatrix} pu(w_0 - t_a) + \alpha u(w_0 - t_b) + qu(w_1 - t_c) - (p+\alpha)u(w_0) - qu(w_1) \\ (1-p-q)u'(w_0 - t_a) - \alpha u'(w_0 - t_b) \\ u'(w_0 - t_a) - u'(w_1 - t_c) \end{pmatrix}.$

Thus equation system (3.15)-(3.17) is equivalent to $\begin{pmatrix} f(t_a,t_b,t_c,\alpha) \\ g(t_a,t_b,t_c,\alpha) \\ h(t_a,t_b,t_c,\alpha) \end{pmatrix} = 0.$

Let $s(\alpha) \equiv \begin{pmatrix} t_a \\ t_b \\ t_c \end{pmatrix}$ be the solution of problem (3.14).

By implicit function theorem, we have

$$D_\alpha s(\alpha) = \begin{pmatrix} f_{t_a}(t_a,t_b,t_c,\alpha) & f_{t_b}(t_a,t_b,t_c,\alpha) & f_{t_c}(t_a,t_b,t_c,\alpha) \\ g_{t_a}(t_a,t_b,t_c,\alpha) & g_{t_b}(t_a,t_b,t_c,\alpha) & g_{t_c}(t_a,t_b,t_c,\alpha) \\ h_{t_a}(t_a,t_b,t_c,\alpha) & h_{t_b}(t_a,t_b,t_c,\alpha) & h_{t_c}(t_a,t_b,t_c,\alpha) \end{pmatrix}^{-1} \begin{pmatrix} f_\alpha(t_a,t_b,t_c,\alpha) \\ g_\alpha(t_a,t_b,t_c,\alpha) \\ h_\alpha(t_a,t_b,t_c,\alpha) \end{pmatrix}.$$

Now we use the following abbreviated notations. Let $a \equiv u'(w_0 - t_a)$, $b \equiv u'(w_0 - t_b)$, $c \equiv u'(w_1 - t_c)$, $x \equiv u''(w_0 - t_a)$, $y \equiv u''(w_0 - t_b)$, $z \equiv u''(w_1 - t_c)$, $u \equiv u(w_0 - t_b)$ and $v \equiv u(w_0)$. Hence, we get

$$\begin{aligned} D_\alpha s(\alpha) &= \begin{pmatrix} -pa & -\alpha b & -qc \\ -(1-p-q)x & \alpha y & 0 \\ -x & 0 & z \end{pmatrix}^{-1} \begin{pmatrix} u-v \\ -b \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{-b^2 z + yz(u-v)}{bxz+apyz-bpxz-bqxz+cqxy} \\ \frac{b(apz+cqx)+(u-v)xz(1-p-q)}{bxz\alpha+apyz\alpha-bpxz\alpha-bqxz\alpha+cqxy\alpha} \\ \frac{-b^2 x + xy(u-v)}{bxz+apyz-bpxz-bqxz+cqxy} \end{pmatrix}. \end{aligned}$$

Since the profit $\pi \equiv pt_a + (1-p-q)t_b + qt_c$, we have

$$
\begin{aligned}
D_\alpha \pi &= pD_\alpha t_a + (1-p-q)D_\alpha t_b + qD_\alpha t_c \\
&= -\frac{1}{(bxz + apyz - bpxz - bqxz + cqxy)\,\alpha}(vxz - uxz - abpz - bcqx + 2puxz \\
&- 2pvxz + 2quxz - 2qvxz + abpqz + bcpqx - 2pquxz + 2pqvxz - quxy\alpha \\
&- puyz\alpha + qvxy\alpha + pvyz\alpha + abp^2z + bcq^2x - p^2uxz + p^2vxz - q^2uxz + q^2vxz \\
&+ b^2qx\alpha + b^2pz\alpha).
\end{aligned}
$$

By (3.17), we have $a = c$ and $x = z$. Replacing c and z by a and x respectively, we get

$$
\begin{aligned}
D_\alpha \pi = &-\frac{1}{(bx(1-p-q) + ay(p+q))\,\alpha}(vx - ux - abp - abq + 2pux - 2pvx \\
&+ 2qux - 2qvx + 2abpq - 2pqux + 2pqvx - puy\alpha + pvy\alpha - quy\alpha + qvy\alpha \\
&+ abp^2 + abq^2 - p^2ux + p^2vx - q^2ux + q^2vx + b^2p\alpha + b^2q\alpha).
\end{aligned}
$$

Since $x, y < 0$ and $a, b > 0$, we have $-\frac{1}{(bx(1-p-q)+ay(p+q))\alpha} > 0$.

By (3.16), we have $b = \frac{1-p-q}{\alpha}a$. Substituting for b, we get

$$
\begin{aligned}
&(vx - ux - abp - abq + 2pux - 2pvx \\
&+ 2qux - 2qvx + 2abpq - 2pqux + 2pqvx - puy\alpha + pvy\alpha - quy\alpha + qvy\alpha \\
&+ abp^2 + abq^2 - p^2ux + p^2vx - q^2ux + q^2vx + b^2p\alpha + b^2q\alpha) \\
&= (v-u)(py\alpha + qy\alpha + x\,(p+q-1)^2).
\end{aligned}
$$

Since x, $y < 0$, we have $py\alpha + qy\alpha + x\,(p+q-1)^2 < 0$. Thus we obtain that $D_\alpha \pi > 0$ if and only if $u > v$.

By lemma 3.4, we have $D_\alpha \pi > 0$ if and only if $\alpha > \underline{\alpha}$. Thus π is increasing in α when $\alpha > \underline{\alpha}$ and decreasing in α when $\alpha < \underline{\alpha}$.

□

A.1.7 Proof of Proposition 3.4

It is clear that the contract in case 1 is a special contract in case 2 when $\alpha = 1-p-q$, thus the profit in case 1 is a constant which is independent of α. By lemma 3.5, the insurer's profit in case 2 is increasing in α when $\alpha > \underline{\alpha}$ and decreasing in α when $\alpha < \underline{\alpha}$. Moreover, by lemma 3.3, we know $\underline{\alpha} < 1-p-q$. Since α is non-negative, we have that when $\alpha \in [\alpha^*, 1-p-q]$ for some $\alpha^* < \underline{\alpha}$, the contract in case 1 is more profitable for the insurer.

□

A.1.8 The Insurance Problem under Justifiability Constraint

Under the justifiability constraint, there must be a full insurance result ($w_0 - t_a = w_0 - t_b = w_1 - t_c$), and additionally (3.15) is satisfied. The insurer's profit is therefore

$$(1-q)t_a + q(w_1 - w_0 + t_a)$$

where t_a is characterized by

$$u(w_0 - t_a) = \frac{p+\alpha}{p+\alpha+q}u(w_0) + \frac{q}{p+\alpha+q}u(w_1).$$

It is clear that the insurer's profit is increasing in α. In addition, if $\alpha = 1-p-q$, the insurer's profit is the same as the profit for the optimal non-vague contract. Figure A.1 depicts the insurer's profit as a function of α.

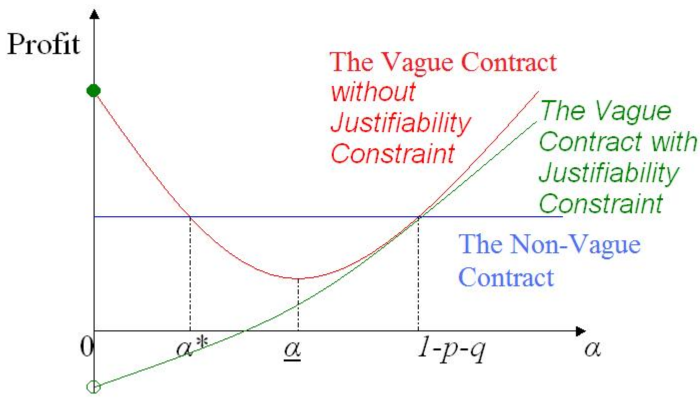


Figure A.1: The profit curves for different α

Note that the profit curve with the justifiability constraint is weakly below the profit curve without it because of this additional constraint for the insurer.

However, when $\alpha = 0$, the insuree is completely unaware of the contingency b. Then every optimal contract without justifiability constraint is justifiable, because the insuree is equally insured at the contingencies, which she is aware of. Thus the profit for the vague contract with justifiability constraint is not continuous at $\alpha = 0$. We therefore conclude that under the constraint of contractual justifiability the insurer will announce b in the optimal contract if and only if $\alpha \in (0, 1-p-q]$.
□

A.1.9 Proof of Proposition 3.5

Let $\begin{pmatrix} f(t,p_1,p_2,\alpha) \\ g(t,p_1,p_2,\alpha) \\ h(t,p_1,p_2,\alpha) \end{pmatrix} \equiv \begin{pmatrix} c'(t^*) - q \\ (1-q)u'(w+t^*-p_1^*-c(t^*)) - \alpha u'(w-p_2^*-c(t^*)) \\ qu(w+t^*-p_1^*-c(t^*)) + \alpha u(w-p_2^*-c(t^*)) - (q+\alpha)u(w) \end{pmatrix}$

$= 0$ and $s(\alpha) \equiv \begin{pmatrix} t^* \\ p_1^* \\ p_2^* \end{pmatrix}$ be the solution.

Now we abbreviate notations. Let $c \equiv c'(t^*)$, $k \equiv c''(t^*)$, $a \equiv u'(w + t^* - p_1^* - c(t^*))(1 - c'(t^*))$, $b \equiv u''(w + t^* - p_1^* - c(t^*))$, $x \equiv u'(w - p_2^* - c(t^*))$, $y \equiv u''(w - p_2^* - c(t^*))$, $u \equiv u(w - p_2^* - c(t^*))$ and $v \equiv u(w)$.

By implicit function theorem, we have

$$\begin{aligned}
D_\alpha s(\alpha) &= \begin{pmatrix} k & 0 & 0 \\ (1-q)b(1-c) - \alpha y(-c) & -(1-q)b & \alpha y \\ qa(1-c) + \alpha x(-c) & -qa & -\alpha x \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ -x \\ u - v \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \frac{x^2 - y(u-v)}{-bx - aqy + bqx} \\ \frac{-aqx - (u-v)(b-bq)}{-bx\alpha - aqy\alpha + bqx\alpha} \end{pmatrix}.
\end{aligned}$$

Since the profit is $\pi^*(\alpha) \equiv qp_1^*(\alpha) + (1-q)p_2^*(\alpha)$, we have

$$\begin{aligned}
D_\alpha \pi^* =& q D_\alpha p_1^* + (1-q) D_\alpha p_2^* \\
=& q(\frac{x^2 - y(u-v)}{-bx - aqy + bqx}) + (1-q)(\frac{-aqx - (u-v)(b-bq)}{-bx\alpha - aqy\alpha + bqx\alpha}) \\
=& \frac{\alpha qvy - \alpha quy + \alpha qx^2 + bv - bu + 2bqu - 2bqv - aqx - bq^2u + bq^2v + aq^2x}{-bx\alpha - aqy\alpha + bqx\alpha}
\end{aligned}$$

Since b, $y < 0$, all other variables are greater than 0, and $q < 1$, we obtain that the denominator $-bx\alpha - aqy\alpha + bqx\alpha = -bx(1-q) - aqy > 0$.

The numerator equals $(v - u)(\alpha qy + b(1-q)^2 + qx(\alpha x - (1-q)a))$. By equation (3.22), we have $\alpha x - (1-q)a = 0$.

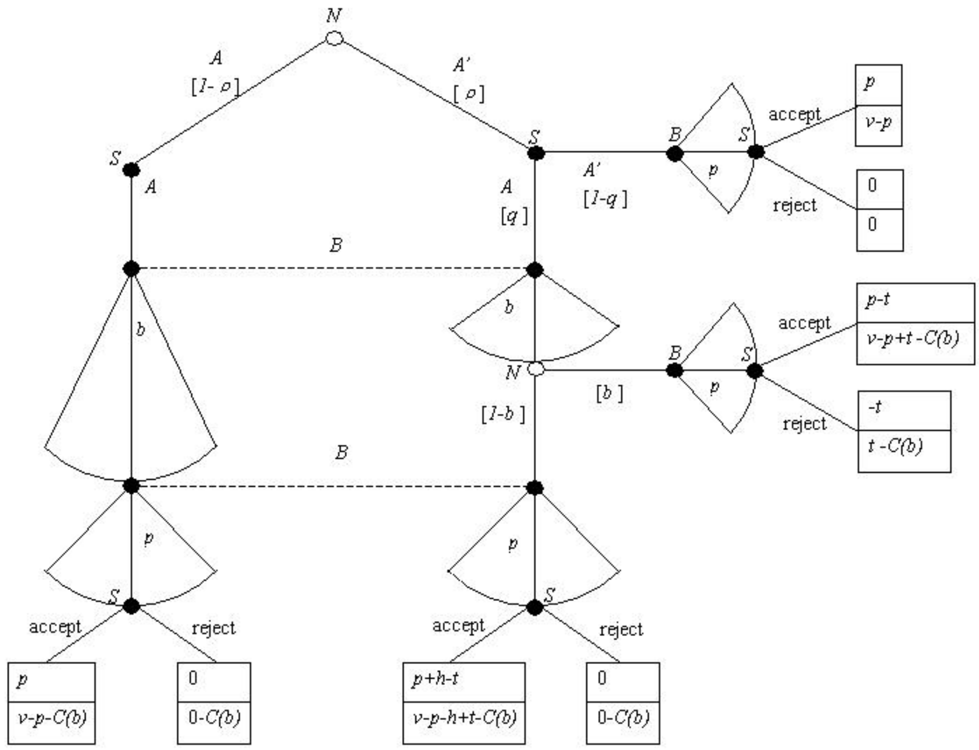
If $\alpha > 1 - q$, then, by equation (3.22), we have $a > x$. Then $u > u'(w + t^* - p_1^* - c(t^*))$. Thus we get $u > v$. We therefore have $(v - u)(\alpha qy + b(1-q)^2) > 0$. Therefore, $D_\alpha \pi^* > 0$. If $\alpha < 1 - q$, the same, we obtain $D_\alpha \pi^* < 0$. Lastly, at $\alpha = 1 - q$, we have $D_\alpha \pi^* = 0$.

Therefore, π gains its minimum at $\alpha = 1 - q$.
□

A.1.10 Proof of Proposition 3.6

Firstly, we show the “if” part: Since $0 < \alpha_g < q$, $\alpha_b > 0$ and $v > 0$, we have $\pi_2 > \pi_1$. Because $0 < \alpha_b < 1 - q$, $\alpha_g > 0$ and $v > 0$, we have $\pi_1 > \pi_3$. Thus $\pi_2 > \pi_3$. Lastly, since $0 < \alpha_b < 1 - q$, $q > 0$ and $v > 0$, we have $\pi_2 > \pi_4$. The contract $C^{\{g\}}$ is therefore optimal for the firm.

Secondly, we show the “only if” part: Since π_2 is the highest profit of all, we have $\pi_2 > \pi_1$ and $\pi_2 > \pi_4$. In addition, $\pi_2 > \pi_1$ implies $\alpha_g < q$, and therefore $\pi_2 > \pi_4$ implies $\alpha_b < 1 - q$.
□



A.1.11 Game Tree

A.1.12 An Infinite-Horizon Bargaining Game where the Buyer makes All the Offers

Instead of assuming a take-it-or-leave-it offer of the buyer in the contracting stage, we consider here a general bargaining protocol where the buyer makes all the offers in an infinite-horizon bargaining game, which is based on Deneckere and Liang (2006). The buyer here can use delay as a screening device to separate the sellers' types.

The buyer's final equilibrium offer must be $p_0 = 0$ where the subscript 0 denotes the last period. It is because any lower offer would not be accepted by type-A seller, whereas any higher offer would be accepted, and therefore dominated. We suppose now there are $n(\geq 0)$ periods that remain before the last period with $p_0 = 0$ in equilibrium. In order to minimize information rent for type-A' seller, type-A' seller should be indifferent between accepting the current offer p_n and waiting n more periods to receive p_0, i.e.,

$$h - t + p_n = \delta^n (h - t + p_0)$$

where δ is the discount factor. Hence, the buyer optimally chooses $p_n = (\delta^n - 1)(h - t)$ in the current period. The buyer's expected payoff is therefore

$$\begin{aligned}&(1-\widehat{\rho})\delta^n (v - p_0) + \widehat{\rho}\,(v - h + t - p_n) - C(b)\\ &= (1-\widehat{\rho})\delta^n v + \widehat{\rho}\,[v - \delta^n (h - t)] - C(b).\end{aligned}$$

It is left to determine the optimal periods of delay n. Assumption 3.5 implies

$$(1-\rho)v > \rho\,(h-t)$$

that also implies

$$(1-\widehat{\rho})v > \widehat{\rho}\,(h-t)\,,$$

or equivalently

$$(1-\delta^n)\,(1-\widehat{\rho})v > (1-\delta^n)\,\widehat{\rho}\,(h-t) \text{ for all } n > 0,$$

or, after some manipulations,

$$(1-\widehat{\rho})v + \widehat{\rho}\,(v-(h-t)) - C(b) > (1-\widehat{\rho})\delta^n v + \widehat{\rho}\,[v-\delta^n\,(h-t)] - C(b) \text{ for all } n > 0.$$

Hence $n = 0$ is optimal for the buyer. That is, the buyer optimally chooses $p_0 = 0$ in the first period, which leads to pooling equilibrium.

Note that since we have implicitly assumed perfect commitment of strategies *ex ante*, we have ignored the renegotiation problem after the buyer observes one period of delay. Nevertheless, the pooling equilibrium is per se renegotiation-proof. Hence, the result $p_0 = 0$ in the first period is also robust to the possibility of renegotiation.

A.1.13 Multiple Buyers in the Pooling Equilibrium

We assume that the economy has two buyers (B_1 and B_2). For the exposition of the idea of the results after introducing multiple buyers, we focus only the pooling equilibrium.

Given buyer 2's cognition level b_2, buyer 1 chooses b_1 to maximize

$$\begin{aligned}&(1-\rho)v + \rho\,(b_1 b_2 + b_1\,(1-b_2) + (1-b_1)\,b_2)\left(v+\frac{t}{2}\right)\\&+\rho\,(1-b_1)\,(1-b_2)\left(v-h+\frac{t}{2}\right) - C(b_1).\end{aligned}$$

Here t is maximal transfer from the seller. When mis-selling is detected, the monetary punishment for the seller is returned to the society. Thus, each buyer shares half of the transfer.

Since the problem is symmetric for buyer 2, the equilibrium cognition levels of two buyers are $b_1 = b_2 = b^*$ such that

$$(1-b^*)\rho h = C'(b^*).$$

Thus, the equilibrium cognition level in the pooling equilibrium is lower compared to the single-buyer case, as each buyer can free ride the other buyer's cognition. However, because there are two buyers here, it is still ambiguous whether the total transaction cost is lower or not.

Note that the analysis above assumes away the possibility of collusion between one buyer and the seller. (Tirole, 1986) When one buyer finds mis-selling and the other does not, the informed buyer can be silent on the information of the mis-selling and ask the seller for a secrete transfer up to t. Thus, the unique informed buyer has to be rewarded by the total transfer in a collusion-proof equilibrium.

Therefore buyer 1 chooses b_1 to maximize

$$\begin{aligned}&(1-\rho)v + \rho b_1 b_2\left(v+\frac{t}{2}\right) + \rho b_1\,(1-b_2)\,(v+t)\\&+\rho\,(1-b_1)\,b_2 v + \rho\,(1-b_1)\,(1-b_2)\left(v-h+\frac{t}{2}\right) - C(b_1).\end{aligned}$$

The equilibrium cognition levels of two buyers are $b_1 = b_2 = b^{**}$ such that

$$(1 - b^{**})\rho h + \frac{\rho t}{2} = C'(b^{**}).$$

Each buyer is therefore incentivized to choose a higher cognition level b^{**} compared to the equilibrium cognition level in the collusion-free case.

Note that it is ambiguous whether each buyer's cognition in the collusion-proof equilibrium is higher or lower than that in the pooling equilibrium in the single-buyer case. The buyer is more likely to choose a higher cognition level in the two-buyer case if the problem of collusion dominates the free riding problem, i.e., t is relatively high.[1]

A.1.14 A Litigation Process to Punish type-A' seller and Reward the Buyer

In this section, we show that the transfer t from the seller to the buyer in the case of mis-selling is bounded above by mechanism design approach.

Assume the goal of the benevolent court of law is to transfer t units of money from the seller to the buyer if and only if A' is appropriate and the seller mis-sells. Note that both parties' preferences over the transfer decision are state-independent, i.e., the seller always prefers no transfer, and the buyer always prefers a transfer. Hence, the standard implementation approach cannot provide unique equilibrium that implements the court's goal. However, in contrast to the standard implementation problem, we assume here that type-A' seller can provide hard evidences of the mis-selling behavior, where as type-A seller and the buyer can only make the cheap talk. The litigation process is based on a mechanism by Ben-Porath and Lipman (2008).

The mechanism is as follows. The seller chooses between providing (the evidence of his type) A' or nothing. If A' is provided, the game ends, and the transfer is realized. If nothing is provided, which is interpreted as the seller's claim of his type A, the buyer then chooses between saying A is appropriate and not. The former is interpreted as the buyer's agreement with the seller, whereas the later is interpreted as her challenge to him. If the buyer says A is not appropriate, the game ends, and the transfer is realized. But the buyer is punished by $f_1(> t)$, and the seller is punished by $\varepsilon(> 0)$. If the buyer says A is appropriate, then the seller has another chance to provide the evidence. If the evidence A' is provided, the transfer is realized. Now the buyer is punished by $f_2(> f_1)$, and the seller is rewarded by $r_2(> t)$. If nothing is provided again, the game ends without any transfer of money.

We now prove that this mechanism implements the court's goal.

Suppose A is appropriate. The seller can provide nothing in each stage. If the buyer says A is appropriate, she receives zero. If not, the buyer receives $t - f_1$. Thus, the buyer will not challenge, and the transfer is not made.

Suppose A is not appropriate. In the last stage, the seller receives $r_2 - t$ for providing the evidence and zero otherwise. Thus, the seller will provide the evidence. Taking it into account, the buyer knows that she receives $t - f_2$ for saying that A is appropriate and $t - f_1$ otherwise. Thus, the buyer will challenge the seller's mis-report. Lastly, in the first stage, the seller will provide the evidence A', since $\varepsilon > 0$.

[1]A similar discussion on the free riding and collusion problems in the context of patent challenges with multiple buyers is in Chiou (2007).

Hence, no matter A is appropriate or not, there is a *unique* subgame perfect equilibrium where the court's goal is implemented.

It is worth mentioning that most buyers cannot be punished arbitrarily off the equilibrium path, due to limited liabilities of the parties. Since the mechanism requires that $f_2 > f_1 > t$, the court cannot raise t arbitrarily to deter the mis-selling here.

Although the mechanism is only one possible litigation process, along the lines of Ben-Porath and Lipman (2008), we can further show that there is no perfect information mechanism[2] with punishment bounded below $\frac{t}{2}$ that implements the court's goal. Suppose by contradiction there is one. Suppose A is appropriate. Let (σ_B^*, σ_S^*) be a subgame perfect equilibrium. Then, for all σ_B, when (σ_B, σ_S^*) is played, the probability that the transfer is realized is strictly less than 1/2. To see it, suppose not. For some σ'_B, when (σ'_B, σ_S^*) is played, the probability that the transfer is realized is greater or equal than 1/2. The buyer then has an incentive to deviate to σ'_B, because her expected payoff is strictly positive, since the punishment is strictly less than $\frac{t}{2}$. So (σ_B^*, σ_S^*) is not an equilibrium, a contradiction. Thus, by playing σ_S^*, the seller receives an expected payoff strictly greater than $-t$, since the probability of being punished is less than 1/2, and the punishment is strictly less than $\frac{t}{2}$. Now suppose A is not appropriate. The strategy sets of the buyer and the seller are the same as before. Then the seller can choose σ_S^* to guarantee an expected payoff strictly greater than $-t$. Hence, the transfer cannot be realized when A is not appropriate, a contradiction.

A.1.15 Connection to Games with Unawareness

Suppose that the seller knows the true game as shown in Appendix A.1.11. However, after the seller says that A is appropriate, at either node in the buyer's non-singleton information set, the game tree from the buyer's view is described in Figure A.2, which is almost the same as the true game, except the prior probability ρ and the name of A'. The buyer still does not know whether something will go wrong for A, and believes that, with probability ϕ, A is not appropriate and some novel product X, which she is unaware of, delivers utility v to her. Here X is a "virtual move" by nature in the buyer's subjective game in terms of Halpern and Rego (2006).

The extended solution of the games with unawareness is the same as the corresponding solution in our original game, except that the objective extent of the mis-selling ρ is replaced by the buyer's subjective one ϕ. Thus, from the behavioral point of view, there is no qualitative difference between these two models whenever $\phi > 0$. As the buyer is aware that something may go wrong for A, she believes that $\phi > 0$. Further, the buyer may have sufficient experience of purchasing products. Thus, it is also plausible that the buyer knows the correct prior ρ. However, the problem is that Bayesian approach in our original model lacks the epistemic foundation of unawareness.

A.1.16 The Case where Assumption 3.5 fails

When Assumption 3.5 fails, the buyer contemplates three alternatives.

First, the buyer exerts b^*, and contracts with both types of sellers. The result is described in Proposition 3.8.

Second, the buyer exerts b', and contracts with only type-A' seller. Under this plan, the buyer proposes

[2] A perfect information mechanism is a sequential-move game where the past actions are perfectly observed. Here the actions are providing hard evidences or cheap talks.

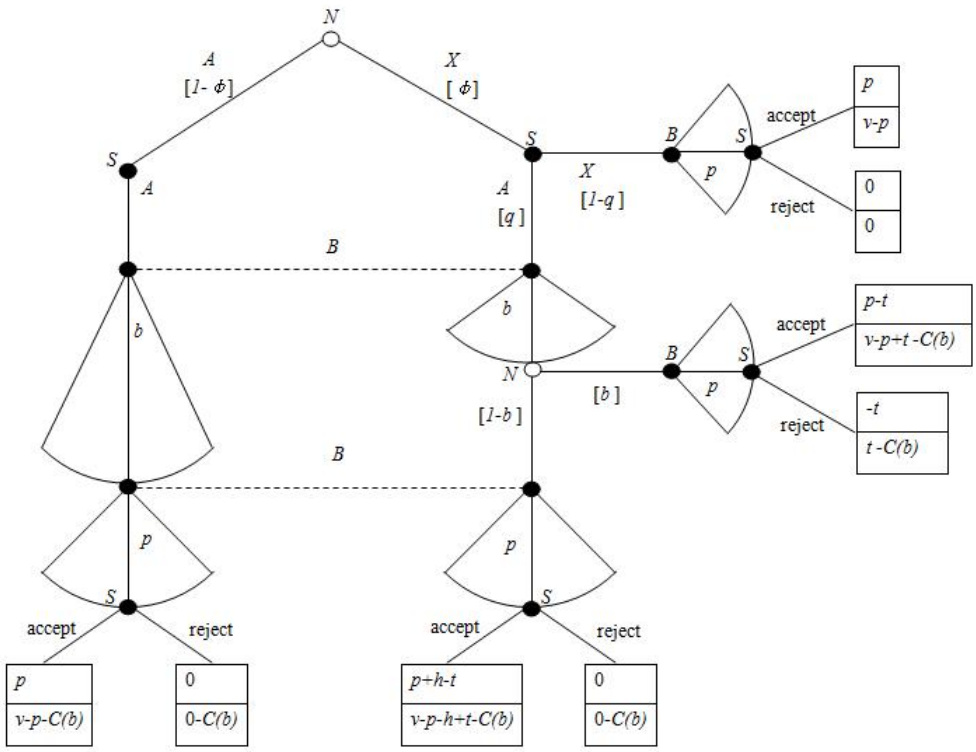


Figure A.2: Game tree from the buyer's view upon observing A

$p = h - t$, and therefore solves the following problem upon observing A:

$$\max_b \frac{(1-\rho)\,0 + \rho q b\,(v+t) + \rho q(1-b)v}{1-\rho+\rho q} - C(b).$$

Compared to the first plan, the buyer loses utility v form contracting with type-A seller, but gains additional utility $h-t$ from type-A' seller in the case where mis-selling is not detected. Since the price $h-t$ is so low that it is common knowledge that the trade occurs only if A is not appropriate, there is no transfer *ex post* in this case.

The optimal cognition level is b' is characterized by

$$C'(b') = \frac{\rho q t}{1-\rho+\rho q}. \tag{A.5}$$

Suppose the second plan is strictly better than the first one for the buyer, which is possible only for a positive q. Equation (A.5) implies b' is also positive. However, because type-A' seller's expected net payoff from mis-selling is

$$(1-b')0 + b'(-t) < 0$$

for any $b' > 0$, type-A' seller's optimal q is zero, a contradiction.

Hence, there is no equilibrium in which the buyer strictly prefers the second plan.

Third, the buyer may exert some cognition effort, and contracts with no sellers. Under this plan, the buyer proposes any price $p > h - t$, and therefore solves the following problem upon observing A:

$$\max_b \frac{(1-\rho)\,0 + \rho q b\,(v+t) + \rho q(1-b)0}{1-\rho+\rho q} - C(b).$$

Along similar lines of the arguments in the second plan, there is no equilibrium in which the buyer strictly prefers the third plan.

A.1.17 Proof of Proposition 3.9

When $\rho h \leq C'(1-\frac{t}{h})$, there is a pooling equilibrium, so $q = 1$. The transaction cost is therefore $L = C(b)$. By (3.27), we have that

$$\frac{dL}{d\rho} > 0,\ \frac{dL}{dh} > 0,\ \text{and}\ \frac{dL}{dt} = 0.$$

When $\rho h > C'(1-\frac{t}{h})$, there is a semi-separating equilibrium.

First, we show $\frac{dL}{d\rho} < 0$.

To see this, by equation (3.31), we have $\frac{dC(b)}{d\rho} = 0$. Thus

$$\frac{dL}{d\rho} = C(b)\frac{d\,(1-\rho(1-q))}{d\rho}.$$

Equation (3.32) implies

$$\frac{d\rho q}{d\rho} = \frac{d\left(\frac{(1-\rho)C'(1-\frac{t}{h})}{h-C'(1-\frac{t}{h})}\right)}{d\rho} = -\frac{C'(1-\frac{t}{h})}{h-C'(1-\frac{t}{h})}.$$

Since $h - C'(1-\frac{t}{h}) > 0$ (otherwise, $q < 0$ by equation (3.32)), we have $\frac{d\rho q}{d\rho} < 0$. Furthermore, we have

$$\frac{d\,(1-\rho(1-q))}{d\rho} = -1 + \frac{d\rho q}{d\rho} < 0.$$

We therefore obtain $\frac{dL}{d\rho} < 0$.

Second, we judge the sign of $\frac{dL}{dh}$.

By definition,

$$\frac{dL}{dh} = (1-\rho(1-q))C'(b)\frac{db}{dh} + C(b)\frac{d\,(1-\rho(1-q))}{dh}.$$

Equation (3.31) and (3.32) imply

$$\begin{aligned}\frac{dL}{dh} &= (1-\rho\,(1-q^*))\,C'(b^*)\frac{t}{h^2} + C(b^*)(1-\rho)\frac{tC''(b^*) - hC'(b^*)}{h\,(h-C'(b^*))^2} \\ &= \frac{1-\rho}{h\,(C'(b^*)-h)^2}\left[tC'(b^*)\,(h-C'(b^*)) + C(b^*)\,(tC''(b^*) - hC'(b^*))\right]\end{aligned}$$

of which the sign is ambiguous. If $tC''(b^*)$ is sufficiently low compared to $hC'(b^*)$, we have $\frac{dL}{dh} < 0$, although we have $h - C'(b^*) > 0$ here. Otherwise, we have $\frac{dL}{dh} > 0$.

Lastly, we show $\frac{dL}{dt} < 0$.

To show this, we first have

$$\frac{dL}{dt} = (1-\rho(1-q))C'(b)\frac{db}{dt} + C(b)\frac{d\left(1-\rho(1-q)\right)}{dt}.$$

By equation (3.31) and (3.32), we have

$$\frac{dL}{dt} = -\frac{(1-\rho(1-q^*))C'(b^*)}{h} - \frac{C(b^*)(1-\rho)C''(b^*)}{h\left(h-C'(b^*)\right)^2} < 0.$$

A.1.18 Proof of zero-rewarding when A is appropriate at the post-contractual stage in Section 3.3.2

We now consider the possibility that a tip is also given to type-A seller.

Consider the following strategy profile in the stage game.

Type-A' seller always unveils A', but the buyer pays some price $\widetilde{p}' > 0$ for A' after A' is reported and pays some price $\widetilde{p} > 0$ for A if A is de facto appropriate at the post contractual stage, and the buyer does not think at all ($b = 0$). The actions in the other nodes in the stage game are the same as in Proposition 3.8.

Detection of deviation in the past leads to the outcome that two parties play the pooling equilibrium in this period.

First, we check type-A' seller's incentive for a deviation. If he deviates from revealing A', his *net* gain in the current period is $h - \widetilde{p}'$.

However, the seller' loss in the future is

$$\frac{\delta}{1-\delta}\left((1-\rho)\,x\widetilde{p} + \rho\left(\widetilde{p}' - (1-x)h - x\,(1-b^*)\,h\right)\right).$$

where b^* is given by equation (3.27). In each period, if A is appropriate, which happens with probability $1-\rho$, the seller loses utility $\widetilde{p}$ given that he is detected. If A is not appropriate, which happens with probability ρ, the seller loses the tip $\widetilde{p}'$ and yet gains the expected rent $(1-x)h + x\,(1-b^*)\,h$ from mis-selling.

Hence, type-A' seller has no incentive to mis-sell if

$$\widetilde{p}' \geq \frac{\left(1-\delta+\delta\rho\,(1-xb^*)\right)h - \widetilde{p}x\delta\,(1-\rho)}{1-\delta+\rho\delta}.$$

Thus, the buyer has to commit to remunerate at least

$$\widetilde{p}'(\widetilde{p}) \equiv \frac{\left(1-\delta+\delta\rho\,(1-xb^*)\right)h - \widetilde{p}x\delta\,(1-\rho)}{1-\delta+\rho\delta}$$

to type-A' seller in order to enhance his truthful report. Note that $\widetilde{p}'$ and $\widetilde{p}$ are *substitutes* for the seller's truthful report.

Second, we check the buyer's incentive for a deviation.

Suppose A is not appropriate. If she deviates from paying $\widetilde{p}'(\widetilde{p})$ for type-A' seller with a truthful report, her net gain in the current period is $\widetilde{p}'(\widetilde{p})$, where the buyer only pays price zero so as to guarantee the seller's acceptance of the contract.

However, the buyer loses her reputation in expectation and the loss in the future is

$$\frac{\delta}{1-\delta}\left(xC(b^*)+\rho x\left(1-b^*\right)h-\rho\widetilde{p}'(\widetilde{p})-(1-\rho)\,\widetilde{p}\right).$$

In each future period, if the buyer's misbehavior is detected, which occurs with probability x, the buyer has to bear the cognition cost $C(b^*)$. With probability $\rho x\left(1-b^*\right)$, she is mis-sold to. But she has not to pay $\widetilde{p}'(\widetilde{p})$ with probability ρ and $\widetilde{p}$ with probability $1-\rho$.

Hence, substituting for $\widetilde{p}'(\widetilde{p})$ and after some manipulation, the buyer has no incentive to misbehave if

$$\widetilde{p}\leq\frac{\delta xC(b^*)-\left(1-\delta\left(1-\rho\left(1-x\right)\right)\right)h}{\delta\left(1-x\right)\left(1-\rho\right)} \tag{A.6}$$

Thus, a smaller $\widetilde{p}$ enhances the buyer's incentive for giving a tip to the seller for his truthful report when A is not appropriate.

Suppose A is appropriate. If she deviates from paying $\widetilde{p}$ for type-A seller at the post-contractual stage, her net gain in the current period is $\widetilde{p}$.

The buyer's loss in the future is also

$$\frac{\delta}{1-\delta}\left(xC(b^*)+\rho x\left(1-b^*\right)h-\rho\widetilde{p}'(\widetilde{p})-(1-\rho)\,\widetilde{p}\right).$$

Similarly, the buyer has no incentive to misbehave if

$$\widetilde{p}\leq\frac{\delta\left((1-\delta+\rho\delta)xC(b^*)+\rho h\left(x+\delta-bx-x\delta-\rho\delta+bx\delta+x\rho\delta-1\right)\right)}{1-\delta+\delta^2\rho\left(1-x\right)\left(1-\rho\right)} \tag{A.7}$$

Hence, by inequation (A.6) and (A.7), we find that $\widetilde{p}=0$ makes separating equilibrium in each period most likely to occur.

Furthermore, when $\widetilde{p}=0$, condition (A.6), which turns to be equivalent to condition (3.34), implies condition (A.7).

To show it, condition (A.7) is equivalent to

$$(1-\delta+\rho\delta)xC(b^*)+\rho h\left(x+\delta-bx-x\delta-\rho\delta+bx\delta+x\rho\delta-1\right)>0,$$

that is,

$$C(b^*)>\frac{\rho h\left(bx-\delta-x+x\delta+\rho\delta-bx\delta-x\rho\delta+1\right)}{(1-\delta+\rho\delta)x}.$$

Therefore, by condition (3.34), it is equivalent to show

$$D\equiv\frac{\left(1-\delta\left(1-\rho\left(1-x\right)\right)\right)h}{\delta x}-\frac{\rho h\left(bx-\delta-x+x\delta+\rho\delta-bx\delta-x\rho\delta+1\right)}{(1-\delta+\rho\delta)x}>0,$$

which always holds, since

$$D=\frac{\left(1-\delta\right)\left(1-\delta+\rho\delta\left(1-bx\right)\right)h}{\delta x\left(1-\delta+\rho\delta\right)}>0.$$

A.1.19 Proof of Proposition 3.15

Obviously, $(A',\,A')$ is an equilibrium, since $t>0$.

If $\frac{h}{2}-t<0$, announcing A' is also a strictly dominant strategy. Thus, (A', A') is the unique equilibrium.

If $\frac{h}{2}-t\geq 0$, by a similar reasoning in section 4.3.2, there are two cases.

The first case is $\rho h \leq C'(1-\frac{2t}{h})$.

In this case, (A, A) is also an equilibrium, because we now have

$$(1-b^*)(\frac{h}{2}-t)+b^*(-t)\geq 0. \tag{A.8}$$

where b^* is given by equation (3.27). For the buyer, after she chooses one of the sellers while facing (A, A), the game is the same as that in section 4.3.2, except that t in the buyer's payoff is doubled, because she may *ex post* sue both type-A' sellers. However, this does not change the equilibrium cognition level b^*, as b^* is independent of t.

We now look for a mixed-strategy equilibrium. Let q_i be the probability that seller S_i chooses A for $i=1$, 2. Since b is determined by q_1 and q_2, in any equilibrium, we have

$$C'(b)=\frac{\rho q_1 q_2 h}{1-\rho+\rho q_1 q_2}. \tag{A.9}$$

Note that if $q_1=0$, then S_2 chooses A' with certainty since $t>0$, which leads to equilibrium (A', A'). If $q_1=1$, then S_2 chooses A with certainty by inequation (A.8), which leads to equilibrium (A, A). Thus, we look for the totally-mixed-strategy equilibrium. Hence, two sellers must be both indifferent between A and A'. The following two conditions therefore hold.

$$q_1\left[(1-b)(\frac{h}{2}-t)+b(-t)\right]-(1-q_1)\,t=0. \tag{A.10}$$

$$q_2\left[(1-b)(\frac{h}{2}-t)+b(-t)\right]-(1-q_2)\,t=0. \tag{A.11}$$

So the mixed-strategy equilibrium is characterized by equation (A.9)-(A.11).

Obviously, in equilibrium, we have $q_1=q_2$. Otherwise, equation (A.10)-(A.11) cannot hold. Hence, the disclosure decisions of A' are strategically complementary for two type-A' sellers in this case.

Furthermore, the totally-mixed-strategy equilibrium exists, because of inequation (A.8) and $t>0$.

The second case to consider is $\rho h > C'(1-\frac{2t}{h})$.

Then (A, A) is not an equilibrium because in this case we have

$$(1-b^*)(\frac{h}{2}-t)+b^*(-t)<0.$$

Furthermore, the totally-mixed-strategy equilibrium does not exist as well, since inequation (A.8) fails and yet $t>0$.

A.1.20 Rationalization of the trap-filtered solution (with cognition)

In this appendix, we provide a microfoundation to rationalize the trap-filtered solution with cognition. Note that this automatically applies to the trap-filtered solution, as it can be seen as a special case in which cognitive thinking is infinitely costly.

Let us consider the following scenario. Suppose that from the agent's viewpoint, there are two types of principals: a normal one (with probability one) and a "crazy" one (with probability zero). The crazy principal always makes a mistake (i.e., offering a non-justifiable contract). However, a normal principal who is rational may intentionally set up a trap for the agent. There are also two types of games: the one where the agent knows the game (with probability one) and the one the agent is unaware of something (and thus does not know the actual game) (with probability zero). The agent is uncertain of the principal's type and her knowledge of the game, and the values of these two variables are independent. Here, we use the *lexicographic probabilistic system* by Blume et al. (1991). First, the event that the principal is normal (crazy) is first-order, and the event of having a crazy principal is second-order. Second, the event that the agent knows the game is first-order, and the event that the agent is unaware of the game is second-order. Further, the event where the principal is crazy and the event in which the agent is unaware of something are of the same order.

Formally, the state space consists of four states, ω_1, ω_2, ω_3, and ω_4 that are differentiated by whether the principal is normal or crazy and whether the agent is fully aware of the entire game. Specifically, let ω_1 represent the state in which the principal is normal and the agent knows the entire game; ω_2 represents the state in which the principal is normal and the agent is unaware of something; ω_3 represents the state in which the principal is crazy and the agent knows the entire game; finally, ω_4 represents the state in which the principal is crazy and the agent is unaware of something. Given the four states, the lexicographic probabilistic system $\mu = (p_1, p_2)$ is as follows: $p_1(\omega_1) = 1$; $p_1(\omega_2) = p_1(\omega_3) = 0$; and $p_2(\omega_2) = \rho$; $p_2(\omega_3) = 1 - \rho$. Put differently, in the terminology of Blume (1991), we have assumed that $\omega_1 >_\mu \omega_2$ and $\omega_1 >_\mu \omega_3$.

Thus, if the agent faces a justifiable contract, she believes that with probability one she knows the game and the principal is normal. However, conditional on a non-justifiable contract, the agent believes that there is a trap with probability ρ and it results from the principal's mistake with probability $1 - \rho$. Therefore, the agent's optimal behavior is exactly as described in the text.

A.1.21 Proof of Proposition 4.2

Proposition 4.2 (with explicit expressions for the optimal contracts)

Let

$$\underline{\tau} = \underline{\tau}(\lambda) = \frac{3(1-\lambda)(\sigma^2+1)+1+\lambda}{(\sigma^2+2)(\lambda+2(1-\lambda)(\sigma^2+1))}$$
$$\overline{\tau} = \overline{\tau}(\lambda) = \frac{(1-\lambda)(\sigma^2+1)+3\lambda-1}{\lambda(\sigma^2+2)}$$

We have $\frac{d}{d\lambda}\underline{\tau} > 0$, $\frac{d}{d\lambda}\overline{\tau} < 0$ for all $\lambda \in [0,1]$, $\underline{\tau}(0) < \tau_{\min} < \tau_{\max} < \overline{\tau}(0)$ and $\underline{\tau}(1) = \overline{\tau}(1) = t^A_{2F}$ The solution of the problem (4.22)-(ICU) is unique and given as follows:

1. If $\tau < \underline{\tau}$ or $\tau > \overline{\tau}$ the incentive constraint (ICU) is slack and the solution is separating, with

$$\alpha^A = \frac{2}{2+\sigma^2}, \qquad \beta^A = \frac{4\left(\sigma^2-2\right)}{2(2+\sigma^2)^2} + \frac{1}{2}(1-\lambda)^2 \frac{\left(1-(\sigma^2+1)\tau-1\right)^2}{\left(1+\sigma^2(1-\lambda)\right)^2},$$
$$\alpha^U = \frac{1+\lambda(\tau-1)}{1+\sigma^2(1-\lambda)}, \quad \beta^U = \frac{1}{2}\tau^2 - \tau\frac{1+\lambda(\tau-1)}{1+\sigma^2(1-\lambda)} - \frac{\left(1-\sigma^2\right)\left(1+\lambda(\tau-1)\right)^2}{2\left(1+\sigma^2(1-\lambda)\right)^2}.$$

2. If $\underline{\tau} \leq \tau \leq \frac{2}{2+\sigma^2}$, the incentive constraint (ICU) is binding, with $\alpha^A - \tau = \tau - \alpha^U$ and

$$\begin{aligned}
\alpha^A &= \frac{1}{1+2\lambda+\sigma^2}\left(2\tau(1+\sigma^2)(1-\lambda) - 1 + 3\lambda + \tau\lambda\right),\\
\beta^A &= \frac{1}{2}\left[\frac{1}{1+2\lambda+\sigma^2}\left(1-3\lambda+(5+2\sigma^2)\tau\lambda\right)\right]^2 - \frac{\tau}{1+2\lambda+\sigma^2}\left(1-3\lambda+(5+2\sigma^2)\tau\lambda\right)\\
&+ \frac{\tau^2}{2} - \frac{1}{2}(2-\sigma^2)\left[\frac{1}{1+2\lambda+\sigma^2}\left(2\tau(1+\sigma^2)(1-\lambda)-1+3\lambda+\tau\lambda\right)\right]^2,\\
\alpha^U &= 2\tau - \alpha^A = \frac{1}{1+2\lambda+\sigma^2}\left(1-3\lambda+(5+2\sigma^2)\tau\lambda\right),\\
\beta^U &= \frac{1}{2}\tau^2 - \frac{1}{2}(1-\sigma^2)\left[\frac{1}{1+2\lambda+\sigma^2}\left(1-3\lambda+(5+2\sigma^2)\tau\lambda\right)\right]^2\\
&- \frac{\tau}{1+2\lambda+\sigma^2}\left(1-3\lambda+(5+2\sigma^2)\tau\lambda\right).
\end{aligned}$$

3. If $\frac{2}{2+\sigma^2} \leq \tau \leq \overline{\tau}$ the solution is pooling, with

$$\begin{aligned}
\alpha^A = \alpha^U &= \frac{1+\lambda+\tau\lambda}{1+2\lambda+\sigma^2},\\
\beta^A = \beta^U &= \frac{1}{2}\tau^2 - \tau\frac{1+\lambda+\tau\lambda}{1+2\lambda+\sigma^2} - \frac{1}{2}\frac{(1+\lambda+\tau\lambda)^2}{(1+2\lambda+\sigma^2)^2}\left(1-\sigma^2\right).
\end{aligned}$$

Proof A.1 *Using Lemmas 4.2 and 4.3, one can eliminate the fixed payment β from the problem and express the contracting problem solely in terms of the incentive component α:*

$$\max_{\alpha^A,\alpha^U} \lambda\left[4\alpha^A - (\sigma^2+2)(\alpha^A)^2 + 2\tau\alpha^U - (\alpha^U)^2\right] + (1-\lambda)\left[2\alpha^U - (\sigma^2+1)(\alpha^U)^2\right] \tag{A.12}$$

$$s.t.\ (\alpha^A-\tau)^2 \geq (\alpha^U-\tau)^2 \tag{A.13}$$

By straightforward differentiation, the unconstrained solution to the maximization problem (A.12)-(A.13) is

$$\alpha^A = \frac{2}{2+\sigma^2}, \quad \alpha^U = \frac{1+\lambda(\tau-1)}{1+\sigma^2(1-\lambda)} \tag{A.14}$$

This solution satisfies the constraint (A.13) strictly if and only if

$$(\tau(\sigma^2+2)-2)^2(\lambda+(1-\lambda)(\sigma^2+1))^2 > (1-\lambda)^2(\tau(\sigma^2+1)-1)^2(\sigma^2+2)^2$$

Viewed as a quadratic inequality in τ, this is equivalent to $\tau < \underline{\tau}$ or $\tau > \overline{\tau}$. Hence, (A.14) yields the separating solution of the proposition.

If (ICU) is binding, there are two possibilities: $\alpha^U = \alpha^A$ (pooling) or $\alpha^U + \alpha^A = 2\tau$ (constrained separating). Direct comparison shows that when $\frac{2}{2+\sigma^2} \leq \tau \leq \overline{\tau}(\lambda)$ we have the pooling solution, and when $\underline{\tau}(\lambda) \leq \tau \leq \frac{2}{2+\sigma^2}$ we have the constrained separating solution.

The monotonicity of $\underline{\tau}(\lambda)$ and of $\overline{\tau}(\lambda)$ follows by differentiation, and the statements about $\underline{\tau}(0), \overline{\tau}(0), \underline{\tau}(1)$, and $\overline{\tau}(1)$ by direct computation.

A.1.22 Proof of Proposition 4.3

Proof A.2 *Firstly, we analyze the case 1 in Proposition 4.2. In this case, we plug the solution under separating solution into the objective function. The profit for the Principal is*

$\pi^S \equiv \lambda\left(\frac{4}{\sigma^2+2} - \frac{8}{(\sigma^2+2)^2} - 2\frac{\sigma^2-2}{(\sigma^2+2)^2} - \frac{1}{2}(\lambda-1)^2\frac{\left(\tau+\sigma^2\tau-1\right)^2}{(\sigma^2\lambda-\sigma^2-1)^2}\right)$

$+ (1-\lambda)\,(\tau - \frac{1}{2}\tau^2 + \tau\frac{\lambda-\lambda\tau-1}{\sigma^2\lambda-\sigma^2-1} - \frac{(\lambda\tau-\lambda+1)^2}{(\lambda+(1-\lambda)(\sigma^2+1))^2}$

$+ (1-\tau)\,\frac{\lambda\tau-\lambda+1}{\lambda+(1-\lambda)(\sigma^2+1)} + \frac{1}{2}(\lambda-\lambda\tau-1)^2\,\frac{1-\sigma^2}{(\sigma^2\lambda-\sigma^2-1)^2}).$

On the other hand, if the Principal makes all Agents aware, he earns a profit $\pi^A \equiv \frac{2}{2+\sigma^2}$.

We get $\pi^A - \pi^S = \frac{(\lambda-1)\left(2\lambda-2\tau-2\lambda\tau+3\sigma^2-3\sigma^2\lambda-3\sigma^2\tau-\sigma^4\tau+3\sigma^2\lambda\tau+2\sigma^4\lambda\tau+2\right)}{(\sigma^2+2)(\sigma^2\lambda-\sigma^2-1)}$.

We further gain that $\pi^A > \pi^S$ *if and only if*

$\tau > R_1 \equiv \frac{-\left(2+\sigma^2\right)\left(\sigma^2\lambda-\sigma^2-\lambda-1\right)+\sigma^2\sqrt{(2+\sigma^2)(\lambda-1)(-\sigma^2+\sigma^2\lambda-1)}}{(2+\sigma^2)(\lambda+\sigma^2+1)}$

or $\tau < L_1 \equiv \frac{-\left(2+\sigma^2\right)\left(\sigma^2\lambda-\sigma^2-\lambda-1\right)-\sigma^2\sqrt{(2+\sigma^2)(\lambda-1)(-\sigma^2+\sigma^2\lambda-1)}}{(2+\sigma^2)(\lambda+\sigma^2+1)}$.

Now we want to check if R_1 *lies to the left or to the right of the boundary of this case. Thus we check the sign of the following term*

$R_1 - \overline{\tau} = \frac{\sigma^2\lambda\sqrt{(\lambda-1)(\sigma^2+2)(\sigma^2\lambda-\sigma^2-1)}-\sigma^2(\lambda-1)\left(\sigma^2\lambda-\sigma^2-1\right)}{\lambda(\sigma^2+2)(\lambda+\sigma^2+1)}$

Since $-\sigma^2(\lambda-1)\left(\sigma^2\lambda-\sigma^2-1\right)$ *is negative,* $R_1 > \overline{\tau}$ *if and only if*

$\sigma^4(\lambda-1)^2\left(\sigma^2\lambda-\sigma^2-1\right)^2 - \sigma^4\lambda^2(\lambda-1)\left(\sigma^2+2\right)\left(\sigma^2\lambda-\sigma^2-1\right)$

$= \sigma^4(2\lambda-1)(1-\lambda)\left(\sigma^2\lambda-\sigma^2-1\right)\left(\lambda+\sigma^2+1\right) < 0.$ *That is,* $\lambda > \frac{1}{2}$.

Thus if $\tau < R_1$ *and* $\lambda > \frac{1}{2}$, *the Principal still uses the solution of separating solution. Otherwise the Principal makes all Agents aware.*

Now we want to check if L_1 *lies to the left or to the right of the boundary of this case. Thus we check the sign of the following term*

$L_1 - \underline{\tau}$

$= \frac{\sigma^2\left(3+2\sigma^2\right)(\lambda-1)\left(\sigma^2(\lambda-1)-1\right)+\sigma^2\left(2(\lambda-1)(1+\sigma^2)-\lambda\right)\sqrt{(\lambda-1)(2+\sigma^2)(\sigma^2(\lambda-1)-1)}}{(2+\sigma^2)(\lambda+\sigma^2+1)(-\lambda+2\sigma^2-2\sigma^2\lambda+2)}$. *It is greater than zero if and only if*

$(\left(3+2\sigma^2\right)(\lambda-1)\left(\sigma^2(\lambda-1)-1\right))^2 - \left(2(\lambda-1)(1+\sigma^2)-\lambda\right)^2(\lambda-1)\left(2+\sigma^2\right)\left(\sigma^2(\lambda-1)-1\right)$

$= -(2\lambda-1)(\lambda-1)\left(\sigma^2\lambda-\sigma^2-1\right)\left(\lambda+\sigma^2+1\right) > 0.$ *That is* $\lambda < \frac{1}{2}$.

Thus if $\tau > L_1$ *and* $\lambda > \frac{1}{2}$, *the Principal still uses the solution of separating solution. Otherwise the Principal makes all Agents aware.*

Secondly, we analyze case 3 in Proposition 4.2. In this case, we plug the solution under pooling solution into the objective function. The profit for the Principal is $\pi^P \equiv \frac{2\lambda+\tau+2\lambda\tau+\lambda^2+\sigma^2\tau-\lambda^2\tau-2\sigma^2\lambda\tau+1}{4\lambda+2\sigma^2+2}$

If the Principal makes all Agents aware, he earns a profit $\pi^A \equiv \frac{2}{2+\sigma^2}$.

We get $\pi^A - \pi^P = \frac{2}{2+\sigma^2} - \frac{2\lambda+\tau+2\lambda\tau+\lambda^2+\sigma^2\tau-\lambda^2\tau-2\sigma^2\lambda\tau+1}{4\lambda+2\sigma^2+2}$.

We further gain that $\pi^A > \pi^P$ *if and only if*

$\tau > R_2 \equiv \frac{-\left(2+\sigma^2\right)\left(\lambda^2-\sigma^2-2\lambda+\sigma^2\lambda-1\right)+(1-\lambda)\sigma^2\sqrt{(2+\sigma^2)(2\lambda+\sigma^2+1)}}{(2+\sigma^2)(2\lambda+\sigma^2-\lambda^2+1)}$

or $\tau < L_2 \equiv \frac{-\left(2+\sigma^2\right)\left(\lambda^2-\sigma^2-2\lambda+\sigma^2\lambda-1\right)-(1-\lambda)\sigma^2\sqrt{(2+\sigma^2)(2\lambda+\sigma^2+1)}}{(2+\sigma^2)(2\lambda+\sigma^2-\lambda^2+1)}$.

Because $t_{2F}^A - L_2 = \frac{\sigma^2(1-\lambda)\left(\sqrt{2+4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda}-\sigma^2-\lambda-1\right)}{(2+\sigma^2)(2\lambda+\sigma^2-\lambda^2+1)}$ *and* $2+4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda-(\sigma^2+\lambda+1)^2 = -\left(\lambda^2-\sigma^2-2\lambda-1\right) > 0$, *we gain that* $\tau < L_2$ *is impossible in this case.*

The same as in case 1, now we want to check if R_2 *lies to the left or to the right of the boundary of this case. Thus we check the sign of the following term*

$R_2 - \overline{\tau} = \frac{\sigma^2(1-\lambda)\left(2\lambda^2-\sigma^2-\lambda+\sigma^2\lambda-1+\lambda\sqrt{4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda+2}\right)}{\lambda(\sigma^2+2)(2\lambda+\sigma^2-\lambda^2+1)} < 0$ *if and only if* $(2\lambda^2-\sigma^2-\lambda+\sigma^2\lambda-1)^2 - \lambda^2(4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda+2)$

$= (2\lambda-1)\left(2\lambda+\sigma^2+1\right)\left(\lambda^2-\sigma^2-2\lambda-1\right) > 0$. *That is,* $\lambda < \frac{1}{2}$.

Thus if $\tau > R_2$ *and* $\lambda < \frac{1}{2}$, *the Principal makes all Agents aware. Otherwise the Principal still uses the solution of pooling solution.*

Thirdly, we analyze case 2 in Proposition 4.2. In this case, we plug the solution under constrained separating solution into the objective function. The profit for the Principal is $\pi^C \equiv \frac{1}{2(2\lambda+\sigma^2+1)}(2\tau - 6\lambda + 20\lambda\tau + 9\lambda^2 - \tau^2 + 2\sigma^2\tau - 10\lambda\tau^2 - 18\lambda^2\tau$

$+10\sigma^2\lambda\tau - \sigma^2\tau^2 + 9\lambda^2\tau^2 - 12\sigma^2\lambda\tau^2 - 12\sigma^2\lambda^2\tau - 4\sigma^4\lambda\tau^2 + 12\sigma^2\lambda^2\tau^2 + 4\sigma^4\lambda^2\tau^2 + 1)$.

We get $\pi^A - \pi^C = \frac{1}{2(\sigma^2+2)(2\lambda+\sigma^2+1)}(20\lambda - 4\tau - 40\lambda\tau + 3\sigma^2 - 18\lambda^2 + 6\sigma^2\lambda + 2\tau^2 - 6\sigma^2\tau + 20\lambda\tau^2 + 36\lambda^2\tau - 2\sigma^4\tau - 40\sigma^2\lambda\tau - 10\sigma^4\lambda\tau - 9\sigma^2\lambda^2 + 3\sigma^2\tau^2 - 18\lambda^2\tau^2 + \sigma^4\tau^2 + 34\sigma^2\lambda\tau^2 + 42\sigma^2\lambda^2\tau + 20\sigma^4\lambda\tau^2 + 12\sigma^4\lambda^2\tau + 4\sigma^6\lambda\tau^2 - 33\sigma^2\lambda^2\tau^2 - 20\sigma^4\lambda^2\tau^2 - 4\sigma^6\lambda^2\tau^2 + 2)$.

The coefficient of τ^2 *is* $-\left(\sigma^2+2\right)\left(9\lambda^2-\sigma^2-10\lambda-12\sigma^2\lambda-4\sigma^4\lambda+12\sigma^2\lambda^2+4\sigma^4\lambda^2-1\right)$

Since $9\lambda^2-\sigma^2-10\lambda-12\sigma^2\lambda-4\sigma^4\lambda+12\sigma^2\lambda^2+4\sigma^4\lambda^2-1$

$= 8\lambda^2-8\lambda+12\sigma^2\lambda^2-12\sigma^2\lambda+\lambda^2-2\lambda+4\sigma^4\lambda^2-4\sigma^4\lambda-1-\sigma^2 < 0$, *we further gain that* $\pi^A > \pi^C$ *if and only if*

$\tau > R_4 \equiv \frac{\left(\sigma^2+2\right)\left(9\lambda^2-\sigma^2-10\lambda-5\sigma^2\lambda+6\sigma^2\lambda^2-1\right)-\sigma^2(1-\lambda)\sqrt{(\sigma^2+2)(2\lambda+\sigma^2+1)}}{(\sigma^2+2)(9\lambda^2-\sigma^2-10\lambda-12\sigma^2\lambda-4\sigma^4\lambda+12\sigma^2\lambda^2+4\sigma^4\lambda^2-1)}$

or $\tau < L_4 \equiv \frac{\left(\sigma^2+2\right)\left(9\lambda^2-\sigma^2-10\lambda-5\sigma^2\lambda+6\sigma^2\lambda^2-1\right)+\sigma^2(1-\lambda)\sqrt{(\sigma^2+2)(2\lambda+\sigma^2+1)}}{(\sigma^2+2)(9\lambda^2-\sigma^2-10\lambda-12\sigma^2\lambda-4\sigma^4\lambda+12\sigma^2\lambda^2+4\sigma^4\lambda^2-1)}$.

It is straightforward to show that $\tau > R_4$ *is impossible in case 2(2).*

Now we want to check if L_4 *lies to the left or to the right of the boundary of this case. Thus we check the sign of the following term*

$L_4 - \underline{\tau}$

$= \frac{\sigma^2(\lambda-1)\left(\left(3+2\sigma^2\right)(1-\lambda)\left(2\lambda+\sigma^2+1\right)+\left(\lambda-2\sigma^2+2\sigma^2\lambda-2\right)\sqrt{2+4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda}\right)}{(2+\sigma^2)(-\lambda+2\sigma^2-2\sigma^2\lambda+2)(9\lambda^2-\sigma^2-10\lambda-12\sigma^2\lambda-4\sigma^4\lambda+12\sigma^2\lambda^2+4\sigma^4\lambda^2-1)}$.

Since $9\lambda^2-\sigma^2-10\lambda-12\sigma^2\lambda-4\sigma^4\lambda+12\sigma^2\lambda^2+4\sigma^4\lambda^2-1 < 0$, *we have* $L_4 - \underline{\tau} > 0$ *if and only if*

$\left(\left(3+2\sigma^2\right)(1-\lambda)\left(2\lambda+\sigma^2+1\right)\right)^2 - \left(\lambda-2\sigma^2+2\sigma^2\lambda-2\right)^2(2+4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda) =$

$= (2\lambda-1)\left(2\lambda+\sigma^2+1\right)\left(9\lambda^2-\sigma^2-10\lambda-12\sigma^2\lambda-4\sigma^4\lambda+12\sigma^2\lambda^2+4\sigma^4\lambda^2-1\right) > 0$. *That is,* $\lambda < \frac{1}{2}$.

Thus if $\tau > L_1$ *and* $\lambda < \frac{1}{2}$, *the Principal makes all Agents aware. Otherwise the Principal still uses the solution of constrained separating solution.*

In summary, we have $\tau_L(\lambda) = L_1$ *and* $\tau_R(\lambda) = R_1$ *when* $\lambda \geq \frac{1}{2}$ *and* $\tau_L(\lambda) = L_4$ *and* $\tau_R(\lambda) = R_2$ *when* $\lambda \leq \frac{1}{2}$ *such that the Principal optimally proposes the contracts identified in Proposition 4.2 if*

$\tau \in (\tau_L, \tau_R)$, and makes all Agents aware otherwise.

It is left to show that $\frac{d\tau_L}{d\lambda} > 0$ and $\frac{d\tau_R}{d\lambda} < 0$ with $\tau_L(1) = \tau_R(1) = t^A_{2F}$, $\tau_L(0) = \tau_{\min}$ and $\tau_R(0) = \tau_{\max}$.

First, we show R_1 is decreasing in λ, because

$\frac{d}{d\lambda}(R_1) = -\frac{\sigma^2\left(3-\lambda+5\sigma^2-4\sigma^2\lambda+2\sigma^4-2\sigma^4\lambda+2(1+\sigma^2)\sqrt{3\sigma^2-2\lambda+\sigma^4-5\sigma^2\lambda-2\sigma^4\lambda+2\sigma^2\lambda^2+\sigma^4\lambda^2+2}\right)}{2(\lambda+\sigma^2+1)^2\sqrt{3\sigma^2-2\lambda+\sigma^4-5\sigma^2\lambda-2\sigma^4\lambda+2\sigma^2\lambda^2+\sigma^4\lambda^2+2}} < 0.$

Second, we show L_1 is increasing in λ.

$\frac{d}{d\lambda}(L_1) = -\frac{\sigma^2\left(\lambda-5\sigma^2-2\sigma^4+4\sigma^2\lambda+2\sigma^4\lambda-3+2(1+\sigma^2)\sqrt{3\sigma^2-2\lambda+\sigma^4-5\sigma^2\lambda-2\sigma^4\lambda+2\sigma^2\lambda^2+\sigma^4\lambda^2+2}\right)}{2(\lambda+\sigma^2+1)^2\sqrt{3\sigma^2-2\lambda+\sigma^4-5\sigma^2\lambda-2\sigma^4\lambda+2\sigma^2\lambda^2+\sigma^4\lambda^2+2}}.$

Since $\lambda - 3 + 4\sigma^2\lambda - 5\sigma^2 + 2\sigma^4\lambda - 2\sigma^4 < 0$, $\frac{d}{d\lambda}(L_1) > 0$ if and only if $(\lambda - 3 + 4\sigma^2\lambda - 5\sigma^2 + 2\sigma^4\lambda - 2\sigma^4)^2 - 4(1+\sigma^2)^2(3\sigma^2 - 2\lambda + \sigma^4 - 5\sigma^2\lambda - 2\sigma^4\lambda + 2\sigma^2\lambda^2 + \sigma^4\lambda^2 + 2)$

$= \left(\lambda + \sigma^2 + 1\right)^2 > 0$ which is always true.

Third, we show R_2 is decreasing in λ, because

$\frac{d}{d\lambda}(R_2) = -\frac{\sigma^2\left(3\lambda+3\sigma^2+\sigma^4+\lambda^3+\sigma^2\lambda+\sigma^2\lambda^2+2+(1+\sigma^2+\lambda^2)\sqrt{4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda+2}\right)}{(\lambda^2-\sigma^2-2\lambda-1)^2\sqrt{4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda+2}} < 0.$

Fourth, we show L_4 is increasing in λ.

$\frac{d}{d\lambda}(L_4) = -\frac{\sigma^2\left(\left(11\sigma^2-12\lambda+3\lambda^2+4\sigma^4-20\sigma^2\lambda-8\sigma^4\lambda+8\sigma^2\lambda^2+4\sigma^4\lambda^2+7\right)\sqrt{4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda+2}-A\right)}{(9\lambda^2-\sigma^2-10\lambda-12\sigma^2\lambda-4\sigma^4\lambda+12\sigma^2\lambda^2+4\sigma^4\lambda^2-1)^2\sqrt{4\lambda+3\sigma^2+\sigma^4+2\sigma^2\lambda+2}}$

where $A \equiv 4 - \lambda^2 + \sigma^2\lambda(5-3\lambda) + \lambda^3 + 12\sigma^2\lambda^3 + 8\sigma^4\lambda^2 + 4\sigma^4\lambda^3 + 4\sigma^6\lambda^2 > 0$.

Thus $\frac{d}{d\lambda}(L_4) > 0$.

Therefore, we have $\frac{d\tau_L}{d\lambda} > 0$ and $\frac{d\tau_R}{d\lambda} < 0$.

Obviously, $R_1 = \frac{2}{2+\sigma^2} = t^A_{2F}$ when $\lambda = 1$ and $L_1 = \frac{2}{2+\sigma^2} = t^A_{2F}$ when $\lambda = 1$. Thus $\tau_L(1) = \tau_R(1) = t^A_{2F}$.

$R_2 = \tau_{\max}$ when $\lambda = 0$ and $L_4 = \tau_{\min}$ when $\lambda = 0$. Thus $\tau_L(0) = \tau_{\min}$ and $\tau_R(0) = \tau_{\max}$.

Bibliography

[1] Abreu, D. (1986): "Extremal Equilibria of Oligopolistic Supergames," *Journal of Economic Theory*, 39, 191-225.

[2] Aghion, P. and P. Bolton (1987): "Contracts as a Barrier to Entry," *American Economic Review*, 77(3): 388–401.

[3] Aghion, P. and P. Bolton (1992): "An Incomplete Contracts Approach to Financial Contracting," *Review of Economic Studies*, 59: 473–493.

[4] Aghion, P., D. Fudenberg and R. Holden (2009): "Subgame Perfect Implementation with Almost Perfect Information," mimeo, Harvard University and MIT.

[5] Aghion, P. and J. Tirole (1997): "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105: 1–29.

[6] Ahn, D. and H. Ergin (2007). "Framing Contingencies", mimeo, University of California at Berkeley.

[7] Akerlof, G. "The Market for Lemons: Quality Uncertainty and the Market Mechanism" *Quarterly Journal of Economics* 84(3), 488-500.

[8] Akerlof G. and W. Dickens (1982). "The Economic Consequences of Cognitive Dissonance". *American Economic Review* 72, 307-319.

[9] Anderlini, L. and L. Felli (1994): "Incomplete Written Contracts: Undescribable States of Nature", *Quarterly Journal of Economics*, 109, 1085-1124.

[10] Anderlini, L. and L. Felli (1999): "Incomplete Contracts and Complexity Costs", *Theory and Decision*, 46, 23-50.

[11] Anderlini, L., L. Felli and A. Postlewaite (2006): "Should Courts Always Enforce What Contracting Parties Write?", STICERD Theoretical Economics discussion paper No. TE/06/510, London School of Economics.

[12] Anderson, L. and C. Holt (1997), "Information Cascades in the Laboratory", *American Economic Review*, 87, 847-862.

[13] Arthur, W.B. 1994. 'Inductive Reasoning and Bounded Rationality', *American Economic Review*, 84 (2), 406-411.

[14] Atkinson, A., S. Mckay, S. Collard and E. Kempson (2007): "Levels of Financial Capability in the UK," *Public Money and Management*, Feb, 29-36.

[15] Aumann, R.J.(1976): "Agreeing to Disagree", *Annuals of Statistics*, 1976, 76(4), 1236-1239.

[16] Avenhaus R., B. von Stengel and S. Zamir (2002): "Inspection Games," in: *Handbook of Game Theory*, 3, 1947-1987.

[17] Bagwell, K. (2007): "The Economic Analysis of Advertising," in: M. Armstrong and R. Porter, *Handbook of Industrial Organization*, vol. 3, North Holland.

[18] Battigalli, P. and G. Maggi (2002): "Rigidity, Discretion, and the Costs of Writing Contracts", *American Economic Review*, 92, 798-817.

[19] Becher, S. (2008): "Asymmetric Information in Consumer Contracts: The Challenge that is yet to be met", forthcoming in *American Business Law Journal*.

[20] Benabou, R. and J. Tirole (2002). "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117(3), 871-915.

[21] Benabou R. and J. Tirole (2003). "Intrinsic and Extrinsic Motivation", *Review of Economic Studies*, 70, 489-520.

[22] Benabou R. and J. Tirole (2006). "Incentives and Prosocial Behavior", *American Economic Review*, 96, 1652-1678.

[23] Ben-Porath, E., and B. Lipman (2008): "Implementation and Partial Provability," mimeo, Hebrew University and Boston University.

[24] Bernheim, D. and M. Whinston (1998): "Incomplete Contracts and Strategic Ambiguity," *American Economic Review*, 88, 902-932.

[25] Bester, H. and K. Ritzberger (2001): "Strategic Pricing, Signaling, and Costly Information Acquisition," *International Journal of Industrial Organization*, 19(9), 1347-1361.

[26] Biglaiser (1993): "Middlemen as Expert," *Rand Journal of Economics*, 24, 212-223.

[27] Board, O. and K.S. Chung (2006). "Object-based unawareness", in: *Logic and the Foundations of Game and Decision Theory*, Proceedings of the Seventh Conference, ed. G. Bonanno, W. van der Hoek, M. Woolridge.

[28] Board, O. and K.S. Chung (2008). "Object-based unawareness II: Applications,"mimeo, University of Minnesota.

[29] Bolton, P. and M. Dewatripont (2005): *Contract Theory*, Cambridge, MA: MIT Press.

[30] Bolton, P., and A. Faure-Grimaud (2007): "Satisficing Contracts," mimeo, Columbia Business School and LSE.

[31] Buchanan, J. M. (1979): *What Should Economists Do?*, Indianapolis: Liberty Press.

[32] Camerer, C., T. Ho, and J. Chong (2004) "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics* 119, 861-898

[33] Chen, Y., and X. Zhao (2009): "Contractual Traps," *Proceedings of the 12th conference on Theoretical Aspects of Rationality and Knowledge* (TARK 2009), Heifetz, A. (ed.), 51-60.

[34] Chiou, J. (2007): "The Design of Post-Grant Patent Challenges", mimeo, NBER.

[35] Chung, K and L. Fortnow(2007): "Loopholes," mimeo, University of Minnesota and Northwestern University.

[36] Crawford, V. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431-1451.

[37] Crémer, J., F. Khalil and J. Rochet (1998a): "Strategic Information Gathering before a Contract is Offered," *Journal of Economic Theory*, 81, 163-200.

[38] Crémer, J., F. Khalil and J. Rochet (1998b): "Contracts and Productive Information Gathering," *Games and Economic Behavior*, 25, 174-193.

[39] Daughety, A. and J. Reinganum (1995): "Product Safety: Liability, R&D, and Signaling," *American Economic Review*, 85(5), 1187-1206.

[40] Daughety, A. and J. Reinganum (2008): "Products Liability, Signaling and Disclosure," *Journal of Institutional and Theoretical Economics*, 164(1), 106-126.

[41] Dekel, E. and F. Gul. 1996. "Rationality and Knowledge in Game Theory", in Kreps and Wallis (eds.), *Advances in Economics and Econometrics: Theory and Application*, vol. I, Cambridge: 87¨C172.

[42] Dekel, E., B. Lipman and A. Rustichini(1998a): "Standard State-Space Models Preclude Unawareness", *Econometrica*, 1998, 66(1), 159-173.

[43] Dekel, E., B. Lipman and A. Rustichini(1998b): "Recent Developments in Modeling Unforeseen Contingencies", *European Economic Review*, 1998, 42, 523-542.

[44] Della Vigna S. And U. Malmendier (2004): "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics* 119, 353-402.

[45] Deneckere, R. and M. Liang (2006): "Bargaining with Interdependent Values," *Econometrica*, 74(5), 1309-1364.

[46] Dixit, A and V. Norman (1978): "Advertising and Welfare," *Bell Journal of Economics*, 9(1), 1-17.

[47] Dye, R. (1985): "Costly Contract Contingencies", *International Economic Review*, 1985, 26(1), 233-250.

[48] Eliaz, K. and R. Spiegler (2006): "Contracting with Diversely Naive Agents," *Review of Economic Studies*, 73, 689-714.

[49] Eliaz, K. and R. Spiegler (2008). "Consideration Sets and Competitive Marketing," mimeo, University of Brown and University College London.

[50] Eyster, E. and M. Rabin (2005). "Cursed Equilibrium". *Econometrica* 73, 1623-1672.

[51] Fagin, R. and J. Halpern (1988): "Belief, Awareness, and Limited Reasoning," *Artificial Intelligence*, 34, 39-76.

[52] Falk A. and M. Kosfeld (2006). "The Hidden Costs of Control", *American Economic Review*, 96(5), 1611-1630.

[53] Fehr, E. and K. Schimdt (1999). "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics* 114, (1999), 817-868.

[54] Feinberg, Y.(2005): "Games with Incomplete Awareness", mimeo, Stanford University.

[55] Feinberg, Y. (2009): "Games with Unawareness," mimeo, Stanford University.

[56] Filiz, E. (2008). "Incorporating unawareness into contract theory", mimeo, Columbia University.

[57] Fiske, S. and S. Taylor (2007). *Social Cognition - From Brain to Culture*, McGraw Hill, 2006.

[58] Friedman, M. (1953): "The Methodology of Positive Economics," in Friedman, M. (eds.), *Essays in Positive Economics*, University of Chicago Press, 1953.

[59] Gabaix, X. and D. Laibson (2006). "Shrouded Attributes and Information Suppression in Competitive Markets",*Quarterly Journal of Economics*, 2006.

[60] Galanis, S.(2006): "Unawareness of Theorems", mimeo, University of Rochester.

[61] Geanakopolos J. (1992). "Common knowledge, Bayesian Learning and Market Speculation with Bounded Rationality." *Journal of Economic Perspectives* 6, 58-82.

[62] Ghirardato, P. (2001): "Coping with Ignorance: Unforeseen Contingencies and Non-Additive Uncertainty," *Economic Theory*, 17, 247–276.

[63] Glaeser, E. and C. Sunstein (2009), "Extremism and Social Learning", *Journal of Legal Analysis*, 1(1), 263-324.

[64] Grant, S. and J. Quiggin (2006). "Learning and discovery", Rice University.

[65] Green, E. and R. Porter (1984): "Non-cooperative Collusion under Imperfect Price Information," *Econometrica*, 52, 87-100.

[66] Grossman, M. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24, 461-83.

[67] Grossman, M. and O. Hart (1980): "Disclosure Laws and Takeover Bids," *Journal of Finance*, 35, 323-24.

[68] Grossman, S. and O. Hart 1983. 'An Analysis of the Principal-Agent Problem', *Econometrica*, 51, 7-45.

[69] Grossman, S. and O. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4): 691-719.

[70] Grossman, S. and J. Stiglitz (1980): "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70, 393-408.

[71] Halpern, J. and L.C. Rego (2006). "Extensive Games with Possibly Unaware Players,"in: *Proc. Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, 744-751.

[72] Hart, O. and J. Moore (1990): "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98(6), 1119-1158.

[73] Hart, O. and J. Moore (2008): "Contracts as Reference Points"," *Quarterly Journal of Economics*, 123(1), 1-48.

[74] Hayek, F. (1945): "The Use of Knowledge in Society," *American Economic Review*, 4, 519-530.

[75] Hayek, F.A. 1967. 'Rules, Perception and Intelligibility', In *Studies in Philosophy, Politics and Economics*, Chicago, The University of Chicago Press, 43-65.

[76] Hayek, F.A. 1978. 'The Primacy of the Abstract', In *New Studies in Philosophy, Politics, Economics and the History of Ideas*, Chicago, The University of Chicago Press, 35-49.

[77] Heifetz, A., M. Meier and B.C. Schipper(2006): "Interactive Unawareness", *Journal of Economic Theory*, 2006, 78-94.

[78] Heifetz, A., M. Meier and B.C. Schipper (2009): "Dynamic Unawareness and Rationalizable Behavior," mimeo, University of California, Davis.

[79] Heller, D. and R. Spiegler (2008): "Contradiction as a Form of Contractual Incompleteness", *Economic Journal*, 118, 875-888.

[80] Hermalin, B. (2002): "Adverse Selection, Short-Term Contracting and the Underprovision of On-the-Job Training", *Contributions to Economic Analysis & Policy*, 1(1), article 5.

[81] Hermalin, B. and M. Katz (1993): "Judicial Modification of Contracts between Sophisticated Parties: A More Complete View of Incomplete Contracts and Their Breach," *Journal of Law, Economics and Organization*, 9, 230-255.

[82] Hirshleifer, J. (1971): "The Private and Social Value of Information and the Reward to Incentive Activity," *American Economic Review*, 61, 561-574.

[83] Holmström, B. 1979. 'Moral Hazard and Observability', *Bell Journal of Economics*, 10, 74-91.

[84] Holmström, B. and P. Milgrom 1991. 'Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design', *Journal of Law, Economics and Organization*, 7, 24-52.

[85] Hu, H. and H. W. Stuart, Jr. (2001). "An Epistemic Analysis of the Harsanyi Transformation", *International Journal of Game Theory* 30, 517-525.

[86] Hughes, G.E. and M.J. Cresswell(1996): *A New Introduction to Modal Logic. Routledge*, London

[87] Inderst, R. and M. Ottaviani (2008): "(Mis)selling through Agents," forthcoming in *American Economic Review.*

[88] Jehiel, P. (2005): "Analogy-Based Expectations Equilibrium". *Journal of Economic Theory* 123, 81-104.

[89] Johnson, E. J., J. Hershey, J. Meszaros, and H. Kunreuther (1993). "Framing, Probability Distortions, and Insurance Decisions," *Journal of Risk and Uncertainty*, 7, 35-51.

[90] Kahneman, D. and A. Tversky. (1973). "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, 5, 207-232.

[91] Kalai G., A. Rubinstein and R. Spiegler(2002): "Rationalizing Choice Functions by Multiple Rationales", *Econometrica* 70, 2481-2488.

[92] Korobkin, R. (2003): "Bounded Rationality, Standard Form Contracts, and Unconscionability", *University of Chicago Law Review*, 70, 1203-95.

[93] Kreps, D. M. (1979): "A Representation Theorem for 'Preference for Flexibility'," *Econometrica*, 47, 565–576.

[94] Kreps, D., P. Milgrom J. Roberts and R. Wilson (1982): "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27(2), 245-252.

[95] Kunimoto, T. (2008): "Indescribability and Asymmetric Information at the Contracting Stage," *Economics Letters*, 99, 367-370.

[96] Laffont, J. and D. Martimort (2002): *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press, 2002.

[97] Laibson D. (1997): "Golden Eggs and Hyperbolic Discounting". *Quarterly Journal of Economics* 62, 443-477.

[98] Leland, E. (1979): "Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards," *Journal of Political Economy*, 87(6), 1328-46.

[99] Li, J. (2009). "Information Structures with Unawareness," *Journal of Economic Theory*, 144, 977-993.

[100] Li, J. (2008): "Unawareness and Zero-Probability,"mimeo, University of Pennsylvania.

[101] Li, S., M. Peitz and X. Zhao (2009): "Worried about Adverse Product Effects? Information Disclosure and Consumer Awareness," mimeo, Toulouse School of Economics and University of Mannheim.

[102] Lipman, B. (2003): "Language and Economics", N. Dimitri, M. Basili and I. Gilboa, eds., *Cognitive Processes and Rationality in Economics*, London: Routledge.

[103] Lizzeri, A. (1999): "Information Revelation and Certification Intermediaries," *Rand Journal of Economics*, 30(2), 214-231.

[104] Martimort, D., and S. Piccolo (2008): "The Strategic Value of Quantity Forcing Contracts,' mimeo, Toulouse School of Economics.

[105] Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23-38.

[106] Maskin, E. and J. Tirole (1999). "Unforeseen Contingencies and Incomplete Contracts", Review of Economic Studies, 66, 83-114.

[107] Matthews, S. and A. Postlewaite (1985): "Quality Testing and Disclosure," *Rand Journal of Economics*, 16(3), 328-340.

[108] Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380-391.

[109] Milgrom, P. (1981): "Acquisition and Disclosure of Information Prior to Sale," *Rand Journal of Economics*, 25, 20-36

[110] Milgrom, P. (2008): "What the Seller Won't Tell You: Persuasion and Disclosure in Markets," *Journal of Economic Perspectives*, 22(2), 115-131.

[111] Milgrom, P. and J. Roberts (1986): "Relying on the Information of Interested Parties," *Rand Journal of Economics*, 17(1), 18-32.

[112] Mirrlees, J. 1975. 'The Theory of Moral Hazard and Unobservable Behaviour: Part 1', *Review of Economic Studies* 66, 3-21.

[113] Modica, S. and A. Rustichini (1994). "Awareness and Partitional Information Structure," *Theory and Decision*, 37, 107-124.

[114] Modica, S., Rustichini, A. and J.-M. Tallon (1998): "Unawareness and bankruptcy: A general equilibrium model", *Economic Theory*, 12, 259-292.

[115] Moore, J. and R. Repullo (1988): "Subgame Perfect Implementation," *Econometrica*, 56, 1191-1220.

[116] Mullainathan, S., J. Schwartzstein and A. Shleifer (2008): "Coarse Thinking and Persuasion," *Quarterly Journal of Economics*, 123(2), 577-619.

[117] Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura (1990): "Strategic Information Revelation," *Review of Economic Studies*, 57, 25-47.

[118] O'Donoghue T. and M. Rabin (1999a). "Doing It Now or Later". *American Economic Review* 89, 103-124.

[119] O'Donoghue T. and M. Rabin (1999b). "Incentives for Procrastinators". *Quarterly Journal of Economics* 114, 769-816.

[120] O'Donoghue T. and M. Rabin (2001). "Self Awareness and Self Control". to appear as a chapter in Roy Baumeister, George Loewenstein, and Daniel Read (eds.), Now or Later: *Economic and Psychological Perspectives on Intertemporal Choice*, Russell Sage Foundation Press, forthcoming.

[121] O'Donoghue T. and M. Rabin (2005). "Incentives and Self Control". mimeo, Cornell University and University of California at Berkley.

[122] Ozbay, E. (2008): "Unawareness and Strategic Announcements in Games with Uncertainty", mimeo, New York University.

[123] Piccione, M. and A. Rubinstein (2003). "Modeling the Economic Interaction of Agents with Diverse Abilities to Recognize Equilibrium Patterns," *Journal of European Economic Association*, 1, 212-223.

[124] Polinsky, A. M. and S. Shavell (2006): "Mandatory versus Voluntary Disclosure of Product Risks," Mimeo, Stanford Law School and Harvard Law School.

[125] Posner, R. (2004): "The Law and Economics of Contract Interpretation", *Texas Law Review*, 83, 1581.

[126] Rabin M.(1998): "Psychology and Economics", *Journal of Economic Literature*, 1998, 36, 11-46.

[127] Rasmusen, E. (2001): "Explaining Incomplete Contracts as the Result of Contract-Reading Costs," *Advances in Economic Analysis & Policy*, 1(1), article 2.

[128] Rego, L.C. and J. Halpern (2007). "Generalized solution concepts in games with possibly unaware players", in: *Theoretical Aspects of Rationality and Knowledge*, Proceedings of the 11th Conference (TARK 2007), Dov Samet (Ed.), Presses Universitaires De Louvain, 253-262.

[129] Rottenstreich, Y. and A. Tversky (1997). "Unpacking, Repacking, and Anchoring: Advances in Support Theory," *Psychological Review*, 104, 406-415.

[130] Rubinstein A.(1998): *Modeling Bounded Rationality*, MIT Press.

[131] Rubinstein, A. (2000): *Economics and Language*, Cambridge: Cambridge University Press.

[132] Rubinstein A. (2006). "Comments on Behavioral Economics". in *Advances in Economic Theory* (2005 World Congress of the Econometric Society), Cambridge University Press.

[133] Sainsbury, M. (1990): "Concepts Without Boundaries", Inaugural Lecture, Department of Philosophy, King's College, London; reprinted in R. Keefe and P. Smith (eds.), *Vagueness. A Reader*, Cambridge MA: MIT Press, 1996, pp. 251-264.

[134] Savage, L. J. (1954): "The Foundations of Statistics". Wiley, New York.

[135] Schwartz, A. and R. Scott (2003): "Contract Theory and the Limits of Contract Law", *Yale Law Journal*, 113, 541-619.

[136] Shapiro, C. (1983): "Premiums for High Quality Products as Returns to Reputations," *Quarterly Journal of Economics*, 98(4), 659-79.

[137] Shavell, S. (1994): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380-391.

[138] Shapiro, J. (2006). "A "Memory-Jamming" Theory of Advertising," mimeo, University of Chicago.

[139] Shavell, S. (2004): *Foundations of Economic Analysis of Law*, Harvard University Press, 2004.

[140] Shavell, S. (2006): "On the Writing and the Interpretation of Contracts", *Journal of Law, Economics and Organization*, 22, 289-314.

[141] Simon, H. (1957): "A Bahavioral Model of Rational Choice," in *Models of Man: Social and Rational*, New York: John Wiley and Sons, 1957.

[142] Skiadas, C. (1997): "Subjective Probability under Additive Aggregation of Conditional Preferences," *Journal of Economic Theory*, 76(2), 242-271.

[143] Spiegler, R. (2006). "The Market for Quacks," *Review of Economic Studies*, 73, 1113-1131.

[144] Spier, K. (1992): "Incomplete Contracts and Signaling," *Rand Journal of Economics*, 23, 432–443.

[145] Strotz R. (1955). "Myopia and Inconsistency in Dynamic Utility Maximization". *Review of Economic Studies* 23, 165-180.

[146] Sunstein, C. and R. H. Thaler (2003). “Libertarian Paternalism”. *American Economic Review*, 93, 175-179.

[147] Tirole, J. (1986): “Hierarchies and Bureaucracies: On the Role of Collusion in Organization,” *Journal of Law, Economics, and Organization*, 2, 181-214.

[148] Tirole, J. (1996): “A Theory of Collective Reputations (with applications to the persistence of corruption and to firm quality),” *Review of Economic Studies*, 63, 1-22.

[149] Tirole, J. (1999): “Incomplete Contracts: Where Do We Stand?” *Econometrica*, 67, 741-781.

[150] Tirole, J. (2009). “Cognition and Incomplete Contracts,” *American Economic Review*, 99(1), 265-94.

[151] Tversky, A. and D. Kahneman (1974). “Judgment under Uncertainty: Heuristics and Biases,” *Science*, 185, 1124-1131.

[152] Tversky, A. and D. Kahneman (1983). “Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment,” *Psychological Review*, 90, 293-315.

[153] Tversky, A. and D. J. Koehler (1994). “Support Theory: A Nonextensional Representation of Subjective Probability,” *Psychological Review*, 101, 547-567.

[154] Vanberg, V. 2002. ‘Rational Choice vs. Program-Based Behavior: Alternative Theoretical Approaches and their Relevance for the Study of Institution’, *Rationality and Society* 14, 7-53.

[155] von Thadden, E. (1992). “Optimal Pricing against a Simple Learning Rule,” *Games and Economic Behavior*, 4, 627-649.

[156] von Thadden, E. and X. Zhao (2009). “Incentives for Unaware Agents,”mimeo, University of Mannheim.

[157] Williamson, O. (1985): *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, New York: Free Press, 1985.

[158] Wu, C. 1590. *Journey to the West* (translated by Collinson Fair), Beijing People’s Press, 1955

[159] Zhao, X. (2008). “Moral Hazard with Unawareness,” *Rationality and Society*, 20(4), 471-496.

[160] Zhao, X. (2009a): " Framing Contingencies in Contracts," mimeo, University of Mannheim.

[161] Zhao, X. (2009b): " Strategic Mis-selling and Pre-Contractual Cognition," mimeo, University of Mannheim.

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig angefertigt und mich keiner anderen als der in ihr angegebenen Hilfsmittel bedient zu haben. Insbesondere sind sämtliche Zitate aus anderen Quellen als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, Jan 2010.

Xiaojian Zhao

Lebenslauf

1981	geboren in Tianjin, China
1999-2003	Bachelor of Management, Jinan University, China
2003-2005	M.A. in Economics and Politics, Universität Freiburg, Deutschland
2005-2010	Promotion in der Volkswirtschaftslehre, Universität Mannheim Mitglied von CDSE