Imperfect Contracting*

Pierpaolo Battigalli
Bocconi University
pierpaolo.battigalli@uni-bocconi.it

Giovanni Maggi
Princeton University
maggi@princeton.edu

February 2000,
revised: November 2001

Abstract

In this paper we model imperfect contracting 'from the ground up,' as arising endogenously from the costs of writing contracts. We model these costs by making explicit the language used to describe the environment and the parties' behavior. The optimal contract may exhibit two forms of incompleteness: discretion, meaning that the contract does not specify the parties' behavior with sufficient detail; and rigidity, meaning that the parties' obligations are not sufficiently contingent on the external state. The model sheds light on the determinants of rigidity and discretion in contracts, and yields rich predictions regarding the impact of changes in the exogenous parameters on the degree and form of contract incompleteness. A simple extension of the model offers a theoretical explanation for the existence of legal default rules (as in the U.S. Uniform Commercial Code) that are meant to "fill the gaps" of private contracts.

*We thank Patrick Bolton, Leonardo Felli, Drew Fudenberg, Gene Grossman, Faruk Gul, Andrea Ichino, Pat Kehoe, Bart Lipman, Salvatore Modica, John Moore, Fabrizio Perri, Jean Tirole, Tim Van Zandt seminar participants at the London School of Economics, the Federal Reserve Bank of Minnesota, Bocconi University, the European University Institute, Florida International University, University of Pennsylvania and participants in the conference "Strategic Rationality in Economics" (Sant'Arcangelo di Romagna, Italy) for very useful comments and discussions. Pierpaolo Battigalli thanks the European University Institute and Bocconi University for financial support. Giovanni Maggi thanks the Department of Economics at NYU and the European University Institute, where he was visiting during part of this project.

1. Introduction

It is often argued that contracts are incomplete because it is too costly to describe all the relevant contingencies and the exact behavior of the contracting parties. In his discussion of the causes of contract incompleteness, Jean Tirole (1999) classifies them in three categories: (i) unforeseen contingencies; (ii) costs of writing contracts; and (iii) costs of enforcing contracts. This paper focuses on point (ii) of this list, which according to Tirole is a weak point in the existing literature: "*... many have argued that contingencies are missing because of substantial costs of writing them. While there is no arguing that writing down detailed contracts is very costly, we have no good paradigm in which to apprehend such costs.*" (Tirole, 1999, p.772).

In a similar spirit, Richard Posner (1992, p. 92) emphasizes the relevance of the costs of writing contracts: "*...some contingencies, even though foreseeable in the strong sense that both parties are fully aware that they may occur, are so unlikely to occur that the costs of careful drafting to deal with them might exceed the benefits, when those benefits are discounted by the (low) probability that the contingency will actually occur.*"

Just what types of costs are incurred in writing contracts is open to debate. For the purposes of this paper, we have in mind costs that are, broadly speaking, proportional to the amount of detail in the contract, such as the cost of figuring out the relevant contingencies and obligations, the cost of thinking how to describe them, the cost of time needed to write the contract, and the cost of lawyers.

We have in mind not only written contracts but also informal (oral) contracts. For example, the relationship between a baby-sitter and the child's parents is typically regulated by an informal contract, in which a set of instructions is communicated orally to the baby sitter (and enforced by the threat of firing her). This set of instructions is typically very incomplete. We believe an important reason for this is that oral communication suffers from similar costs of complexity as written contracts. To keep the terminology lean, however, in the remainder of the paper we will simply talk about "writing costs".

Even though the above quotes by Jean Tirole and Richard Posner emphasize the notion of "missing contingencies," it is clear at a moment's reflection that contract incompleteness can take two distinct forms: (excessive) *discretion,* meaning that the contract does not specify the parties' behavior with sufficient precision; and *rigidity*, meaning that the parties' obligations are not sufficiently contingent on the external state. For example, a construction contract is characterized by discretion if it does not specify the materials with sufficient precision (and this results in the contractor choosing low-quality materials); and is characterized by rigidity if the completion time for the project is fixed, when it would be more efficient to make it contingent on certain exogenous events.

The presence of writing costs can explain both of these forms of incompleteness. Intuitively, if it is costly to describe the external contingencies *and the parties' behavior*, then there is a potential role for both rigidity and discretion. In this paper we explore this intuition more rigorously. We will present a model that sheds light on the implications of writing costs for the optimal degrees of rigidity and discretion in contracts, and is tractable enough to generate potentially testable predictions about the impact of changes in the fundamental parameters.

We now sketch the structure of the model and our main results. In section 2, we develop a framework that makes explicit the language used to describe the environment and the parties' behavior. In particular, the language is made of (i) *primitive sentences* describing elementary events and elementary actions, and (ii) logical connectives (such as "not", "and", "or"). This language can be used to describe state-dependent constraints on behavior. Each primitive sentence has a cost (logical connectives have zero cost), and the total cost of writing the contract is a function (e.g. a summation) of the costs of its primitive sentences. It should be emphasized that our simple language is best suited to describe qualitative aspects of the environment and behavior.

We consider a simple principal-agent framework with symmetric information, where parties are risk neutral and contracts are perfectly enforceable. A contract may specify contingent obligations for the agent and a fixed transfer from the principal to the agent. After the contract is signed, the principal makes the agreed-upon transfer to the agent, then the state is realized and observed, then the agent takes the relevant actions.

In section 3, we impose more structure on our framework, in order to obtain more tractability and sharper predictions. We consider a model in which the agent can take a set of dichotomous elementary actions (such as "feed the baby" versus "do not feed the baby") and the state is described by a set of dichotomous elementary events (such as "the baby cries" versus "the baby does not cry"), with a one-to-one correspondence between the two (so that, for example, it is efficient to feed the baby if he cries, and not feed the baby if he does not cry). We label this a "match-the-state" setting. We characterize the structure of the optimal contract and analyze how this changes with the parameters. In the optimal contract, the most important actions are regulated by contingent clauses; a set of less important actions is regulated by rigid rules; and the least important actions are left to the agent's discretion (i.e. they are not regulated at all). Any of these sets may be empty, depending on the parameters.

A key parameter is the writing cost. For moderate levels of the writing cost, the optimal contract displays some rigidity; for higher levels of the writing cost the optimal contract is characterized by both rigidity and discretion; and for very high levels of the writing cost the empty contract is optimal. The same is true if we increase the complexity of the contractual relationship (as measured by the number of elementary actions/events). We also find that the length of the contract is nonmonotonic in the complexity of the contractual relationship, reaching a maximum for intermediate levels of complexity.

Another interesting comparative-statics result concerns the impact of uncertainty. The model predicts that in more uncertain environments contracts should contain more contingent clauses and fewer rigid clauses, and should leave more discretion to the agent. It is interesting to note that, as uncertainty increases, the optimal contract may become simpler, in the sense of a lower total complexity cost.

The tractability of our basic model depends on a number of assumptions on the payoff structure, on the form of contracts and on the language used to write contracts. In section 4 we discuss the robustness of our qualitative results to changes in these assumptions. In section 5 we discuss how results are likely to change in the presence of unforeseen events.

In section 6 we argue that a simple extension of our model can provide a theoretical explanation for the existence of legal default rules (as in the U.S. Uniform Commercial Code), whose purpose is to "fill the gaps" of private contracts. The key idea is that legal default rules can save on complexity costs. This explanation of legal defaults has been proposed by several law-and-economics scholars, though at an informal level. Some of these scholars have also expressed the view that the optimal legal default rules are the ones that the majority of contracting pairs would agree to in the absence of transaction costs (this is the so-called "majoritarian" theory of legal defaults).[1] We will not attempt a thorough examination of this issue in this paper, but will suggest that our framework is a natural one to address this type of question.

At this point, the skeptical reader may still ask: how important are writing costs in reality? This is ultimately an empirical question, that we will not be able to settle here, but we will offer a few remarks and casual observations. A preliminary consideration is that the cost of including one additional clause in a contract may well be small, but for most contracting situations the number of events and actions that are potentially relevant is arguably astronomical, so that the cost of writing a complete contract would be extremely large. The following example should strengthen this point. Consider a principal who delegates the writing of a document to an agent (this could be a lawyer delegating the writing of a letter to an assistant, or a professor delegating the writing of a survey to a research assistant). Of course there is an astronomical number of possible documents that can be written. A complete contract would describe *exactly* the document that the principal wants to see, but this would involve nothing short of writing the whole document, thus defeating the whole purpose of the trade. Instead, it may be optimal to give the assistant an incomplete set of instructions, specifying some general characteristics that the document should have, the number of pages, etc.[2] In this type of situation, writing costs are relevant almost by logical necessity; more generally, the example suggests that, even if the 'unit' writing cost is very small, the total cost of a complete contract is easily blown up by the dimensionality of the contracting problem.

The example we just gave concerns the costs of describing behavior. As for the costs of

[1] For example, Easterbrook and Fischel (1982) propose that "corporate law should contain the defaults most people would have negotiated, were the costs of negotiating at arms' length for every contingency sufficiently low". Goetz and Scott (1983) propose that the state should set default rules by asking "what arrangements would most bargainers prefer?" Similar statements of this theory can be found in Posner (1986) and Baird and Jackson (1985). It must be emphasized that several scholars, for example Ayres and Gertner (1989, 1992), have expressed reservations about this majoritarian view of legal defaults. More on this in section 6.

[2] In this example, the agent does not have better information than the principal, so the incompleteness of instructiond is caused solely by communication costs. Of course, if the agent had superior information there would be an additional reason for giving incomplete instructions. We are abstracting from this type of consideration here.

describing contingencies, it is not hard to find examples of contracts where relevant contingencies are missing even though they are foreseeable and verifiable, thereby suggesting the presence of writing costs. An example is provided by Meihuizen and Wiggins (2000), who examine the evolution of supply contracts in the natural gas industry between 1946 and 1985 in the United States. Around 1975, most of these contracts were amended to include a new clause that provided for renegotiation of the price in case of deregulation of the industry. Our interpretation is that, before it was introduced, this was a classic "missing contingency." Since the industry was regulated, the contracting parties were almost by definition aware of the possibility of deregulation. We are therefore inclined to think that this contingency was missing because it was considered very unlikely, and was later introduced because its likelihood was revised upwards (possibly because of the 1973-74 oil crisis), or more generally because the expected benefit of writing this clause came to exceed its cost.[3]

This is not the first paper that explicitly models the complexity of writing contracts as a source of contractual incompleteness. The pioneering paper in this literature is Dye (1985), and more recent papers include Anderlini and Felli (1994, 1998, 1999), Krasa and Williams (1999) and Al Najjar *et al.* (2001). Before discussing these papers in more detail, we highlight in general terms what we think is our main contribution to this literature. The above-mentioned papers model contracts as functions mapping external states into an outcome (typically a monetary transfer). As a consequence, in these models, contractual incompleteness can only take the form of rigidity. In our framework we consider other contractual obligations besides monetary transfers, and we assume that a detailed description of such contractual obligations is costly; therefore our model is capable of explaining both rigidity and discretion. A related innovation of our model is that it makes explicit the language used to write contracts; this allows a simple and intuitive formalization of the costs of describing the environment and behavior.

[3]Another example of missing contingencies can be found in the area of environmental insurance contracts. Many insurance companies have recently introduced a new contingent clause in their pollution-insurance contracts. This clause excludes injuries caused by (spores released by) certain strains of mold that grow in buildings. In the past, insurance companies had received some claims related to this type of injury, but the frequency of these claims was very low. The frequency of claims for some reason increased substantially in recent times, and as a consequence the new exclusion clause was added to the contracts. We view this anecdote as suggestive of nonnegligible writing costs. If writing additional clauses were costless, probably the exclusion clause on mold would have been introduced from the beginning.

Dye (1985) explains the presence of rigidity by assuming that the cost of writing a contract is increasing in the number of its contingencies, that is, the number of cells in the partition of the state space induced by the contractual function. Our model differs from Dye's in several dimensions. First, we view the complexity of a contract in a very different way. For example, two contracts with the same number of mutually exclusive contingencies have the same cost according to Dye, but could have very different costs in our model.[4] This is because, in our framework, the cost of a contract is not a function of the number of contingencies specified in the contract, but of how hard it is to describe those contingencies in the given language. Second, the two models yield different comparative-statics predictions, as we will discuss in section 2.8. Finally, as already mentioned, our model is able to explain the presence of discretion in contracts, while Dye's model is not (more on this in section 4.5).

Anderlini and Felli (1994) capture the difficulty of describing contingencies in a different way: in a co-insurance model with a continuum of states, they require that contracts correspond to computable functions, i.e., algorithms that for every input (state) produce an outcome in a finite number of steps. They show that the computability constraint *per se* does not preclude an approximate first best. But if the decision process used to select the contract is also constrained to be algorithmic, the resulting contract is incomplete.[5] Krasa and Williams (1999) consider a similar constraint on the complexity of a contract: they assume that the number of relevant contingencies (elementary dummy variables) is countably infinite, but the contractual outcome can depend only on a finite number of contingencies. They explore the conditions under which the optimal contract can be approximated (in a payoff metric) by contracts satisfying this finiteness constraint. Anderlini and Felli (1999) is closer to our work. They consider a large class of complexity measures for computable functions satisfying a few plausible axioms, and show that for any complexity measure in the given class one can find a contracting problem such that the optimal contract is incomplete. Broadly speaking, our approach differs from theirs in that we impose more structure on the problem and in return we get sharper predictions from

[4] Consider the following two contracts: contract A specifies behavior b^0 if the exogenous event E occurs and behavior b^1 otherwise; contract B specifies behavior b^0 if the exogenous events E and F occur and behavior b^1 otherwise. These contracts have the same complexity cost according to Dye's assumption, whereas contract B is more costly according to our model.

[5] Anderlini and Felli (1998) show that the approximation result of Anderlini and Felli (1994) fails when the parties' utilities are discontinuous. Al Najjar *et al.* (2001) present a model with a countable state space, finitely additive probabilities, and continuous utilities, where the approximation result also fails.

the model.[6]

Before plunging into the analysis, we need to comment briefly on the well-known irrelevance result by Maskin and Tirole (1999). They show that the possibility of unforeseen contingencies and the costs of describing contingencies need not imply inefficiencies in contracting, provided a message-based mechanism can be played after the state is observed and before actions are taken. We think our approach is useful in spite of the Maskin-Tirole result. First, in many situations it is not feasible to play games after the state is realized and before actions are taken.[7] Second, even if it is feasible to play a mechanism a' la Maskin-Tirole, it is still necessary to describe *behavior*, which can be quite complicated. Third, a Maskin-Tirole mechanism can itself be quite complex, and the costs of describing and implementing the mechanism might not be lower than those of describing the relevant contingencies. Fourth, as shown in Battigalli and Maggi (2000b), if parties interact repeatedly and can contract at any point in time, writing costs can lead to inefficiencies even if mechanisms a' la Maskin-Tirole are available.[8]

The paper is structured as follows. In section 2 we present our framework. In section 3 we analyze the basic model with conflict of interests and derive the main results. In section 4 we discuss the implications of more general payoffs, of more general contract forms and of richer languages. Section 5 offers some comments on unforeseen events. Section 6 addresses the issue of legal default ruels. Section 7 offers concluding remarks about an extension of our framework.

[6] The literature has pointed out a number of other potential causes of contract incompleteness beside the costs of writing contracts. Allen and Gale (1992), Spier (1992) and Dewatripont and Maskin (1995) argue that the presence of asymmetric information can be a source of contract incompleteness. Boot et al. (1993) argue that an optimal contract may exhibit discretion when some contingencies are not verifiable. The reason is that, if contingencies are not verifiable, a contract that completely specifies behavior would force non-contingent actions, while a contract leaving some discretion may induce the agent to respond efficiently to contingencies. A similar argument is made in Bernheim and Whinston (1998). In their model, the parties take actions in sequence. If the first-mover's actions are not verifiable, the optimal contract may leave some discretion in the second-mover's choice of actions. Mukerji (1998) argues that, if parties are ambiguity-averse rather than expected-utility maximizers, the equilibrium contract may be excessively rigid.

[7] Consider the baby-sitting example: the baby sitter must react quickly to contingencies, and playing a mechanism with the baby's parents before taking action is out of the question.

[8] Further qualifications to the Maskin-Tirole result have been pointed out by Segal (1999) and Hart and Moore (1999). Segal (1999) considers a hold-up problem in which contingencies cannot be described ex-ante, parties cannot commit not to renegotiate, and only a finite number of actions can be described ex-post. He shows that, even if message-contingent mechanisms a' la Maskin-Tirole are available, the first-best outcome cannot be achieved. Moreover, as the size of the action space grows, the benefit from any message-contingent mechanism shrinks (see also Segal, 1995). Hart and Moore (1999) consider a hold-up problem similar to Segal's, and show that the first-best outcome may be unattainable *even if* states can be costlessly described ex-ante.

All the proofs are contained in the appendix.

2. The Framework

2.1. Language

Our starting point is a simple formalization of the language used to write a contract.

$\Pi^e = \{e_1, e_2, e_{3,}...\}$ is a finite collection of *primitive sentences*, each of which describes an *elementary event.* These are the exogenous aspects of the world that are relevant to the contracting problem, for example, e_1: "the baby cries", e_2: "the baby smells", e_3: "it rains". With a slight abuse of terminology, we will use the notation e_k to indicate both an elementary event and the primitive sentence that describes it.

$\Pi^a = \{a_1, a_2, a_3, ..\}$ is a finite collection of primitive sentences describing *elementary actions* (or *tasks*), for example, a_1 : "feed the baby", a_2 : "change diapers", a_3 : "take baby for a walk". The set of all primitive sentences is denoted $\Pi = \Pi^e \cup \Pi^a$. The simple language that we formalize here is best suited to describe *qualitative* aspects of the environment and behavior.

Using the primitive sentences, the logical connectives $\neg$ ("not"), $\wedge$ ("and"), $\vee$ ("or"), $\rightarrow$ ("if... then"), the parentheses and the logical constant $\top$ ("tautology") we can derive well-formed formulae about the exogenous environment and/or about behavior.[9] A formula about the environment describes a contingency, for example $(e_1 \vee e_2) \wedge (\neg e_3)$ ("it does not rain and the baby cries or smells"). A formula about behavior describes a set of instructions, for example $(a_1 \vee a_3)$ ("feed the baby or take him for a walk"). We will use interchangeably the expressions "formula about the environment" and "contingency", and likewise for expressions "formula about behavior" and "set of instructions". The logical constant $\top$ in our setting will be used only in two ways: as a formula about the environment it will mean "whatever happens," and as a formula about behavior it will mean "anything."

The set of well-formed formulae about the environment is denoted Λ^e, the set of well-formed

[9] A formula is "well-formed" if it is constructed according to the rules of the language, which are quite similar to those used in algebra. See, e.g., Hamilton (1988) for details.

formulae about behavior is denoted Λ^a.

An important assumption is that the language just described is the (only) common-knowledge language for the parties and the courts. This ensures that there are no problems of ambiguous interpretation of the contract.

2.2. Contracts

We consider formal contracts between a principal and an agent. A contract stipulates a number of clauses of the form "if contingency η_k occurs then the agent must follow the instruction β_k," and a monetary transfer from the principal to the agent. We will represent a nonmonetary clause as a formula $\eta_k \to \beta_k$, and we will call "job description" a conjunction of such clauses. We could allow transfers to be contingent on the external environment and/or behavior, but there would be no gains from doing so, due to the assumptions (to be introduced shortly) of verifiable states and behavior, risk neutrality and conflict of interests (see section 4.5).

Definition 1. *A* contract *is a job description,* $g = \bigwedge_{k=1}^{K}(\eta_k \to \beta_k)$ $(\eta_k \in \Lambda^e, \beta_k \in \Lambda^a)$*, and a transfer* $t \in \mathbf{R}$.

Examples of contract clauses are: (1) $\neg e_3 \to a_3$, "if it does not rain, take baby for a walk"; (2) $(e_1 \vee e_2) \to (a_1 \wedge a_2)$, "if baby cries or smells, feed him and change his diapers"; (3) $\top \to (a_4 \vee a_5)$, "always talk to the baby or sing to the baby".

Note that the different contingencies η_k, $k = 1, ..., K$, will in general *not* be mutually exclusive, as a contract with mutually exclusive contingencies may be more complex than an equivalent contract with non exclusive contingencies. Similarly, a contingency η_k will in general not be a complete description of the environment and an instruction β_k will in general not be a complete specification of behavior.

Since the transfer will be determined so as to make the agent indifferent between accepting and rejecting the contract, it will play no interesting role in the analysis. For this reason, with a slight abuse of our terminology, we will often refer to the job description g simply as the "contract".

2.3. Costs of Writing Contracts

We assume that all primitive sentences have the same writing cost c. To obtain the cost of contract g we assume that one has to pay the cost of each primitive sentence occurring in the contract plus the cost r of "recalling" a primitive sentence for each occurrence after the first one. We assume that the recalling cost is (weakly) smaller than the cost of writing a primitive sentence: $0 \leq r \leq c$.

We also assume that writing the logical connectives, the logical constant and the transfer has no cost. Let n^g be the number of *distinct* primitive sentences occurring in contract g, and $\bar{n}^g$ the *total* number of primitive sentences contained in contract g. Then the cost of writing g is:[10]

$$C(g) = cn^g + r(\bar{n}^g - n^g). \tag{2.1}$$

For example, if g consists of clauses (1)-(3) in the previous subsection, the cost of writing contract g is $8c$.

2.4. States and Behavior

A *state of the environment* (or simply a *state*) is a complete description of the exogenous environment, represented by a valuation function $s : \Pi^e \to \{0, 1\}$ (that is, $s \in \{0, 1\}^{\Pi^e}$). $s(e_k) = 1$ means that primitive sentence e_k is true at state s and $s(e_k) = 0$ means that primitive sentence e_k is false at state s. In other words, $s(e_k)$ is a dummy variable that takes value one if elementary event e_k occurs and zero otherwise; and a state can be thought of as a realization of the vector of dummy variables $(s(e_1), s(e_2), ...)$.

A simple example can illustrate the relationship between elementary events and states. Suppose the only relevant elementary events are e_1 ("the baby cries") and e_2 ("the baby smells").

[10]Note that our language does not allow for a simple way to express a clause of the form "if none of the contingencies described in the rest of the contract applies, then β_j." This kind of clause can be expressed in our language only by recalling all contingencies η_k described in the rest of the contract, namely as $\left(\neg \left(\bigvee_{k \neq j} \eta_k\right) \to \beta_j\right)$. If the recalling cost r is zero, then our cost formula captures the fact that a clause of this kind can be written at low cost. If $r > 0$, then our cost formula does not properly take this into account. While extending our language to allow for expressions of the kind "if none of the above applies" is conceptually straightforward, it is also notationally cumbersome and it would not alter the qualitative features of the analysis.

Then there are four relevant states, as in the following table:

	e_1	$\neg e_1$
e_2	baby cries and smells	baby smells and does not cry
$\neg e_2$	baby cries and does not smell	baby does not cry and does not smell

Function s is extended on Λ^e in the following inductive way:

- $s(\top) = 1$,
- $s(\neg\eta) = 1$ if and only if $s(\eta) = 0$,
- $s(\eta \vee \epsilon) = \max(s(\eta), s(\epsilon))$,
- $s(\eta \wedge \epsilon) = \min(s(\eta), s(\epsilon))$.

Similarly, a *behavior* is a complete description of what the agent does, represented by a valuation function $b : \Pi^a \to \{0,1\}$ (that is, $b \in \{0,1\}^{\Pi^a}$). $b(a_k) = 1$ means that elementary activity a_k is executed, and $b(a_k) = 0$ that a_k is not executed. The function b is extended on Λ^a analogously as function s. To illustrate the relationship between elementary actions and behaviors, suppose the only elementary actions are a_1 ("feed the baby") and a_2 ("change the baby's diapers"). Then there are four relevant behaviors:[11]

	a_1	$\neg a_1$
a_2	feed and change diapers	change diapers and do not feed
$\neg a_2$	feed and do not change diapers	do not feed and do not change diapers

2.5. The behavioral correspondence

Some states of the environment may be impossible because they contradict the factual knowledge of the parties (e.g. the laws of mechanics) and/or the meaning of the primitive sentences.

[11] In principle the parties could construct a new language, for example by attaching a number to each state and to each behavior, and write a contract with the new, possibly simpler language. *However*, given the assumption that our propositional language is the only common-knowledge language for the parties and the courts, writing the contract in a new language will not help. The reason is that the parties will have to attach a vocabulary that translates the new language in terms of the original, common-knowledge language, in order for the courts to interpret the new language, and doing so is at least as costly as writing the contract in the original language. We thank Leonardo Felli for bringing this issue to our attention.

We let $S \subseteq \{0,1\}^{\Pi^e}$ denote the set of possible states. Likewise, for each state, some behaviors may be impossible. We let $\overline{B}(s) \subseteq \{0,1\}^{\Pi^a}$ denote the set of possible behaviors for any given state $s \in S$.

A contract $g = \bigwedge_{k=1}^{K}(\eta_k \to \beta_k)$ imposes state-dependent constraints on the behavior of the agent. These constraints, together with the feasibility constraints, constitute what we call the *behavioral correspondence* induced by the contract. This behavioral correspondence can be derived logically from the contract g in the following way.

For every contingency η_k, define the *truth set* of η_k, denoted by $\|\eta_k\|$, as the set of possible states where contingency η_k is true, i.e.

$$\|\eta_k\| = \{s \in S : s(\eta_k) = 1\}.$$

and define analogously the truth set of β_k, denoted by $\|\beta_k\|$, as the set of behaviors b such that the instructions β_k are satisfied. The behavioral correspondence induced by g is

$$B^g(s) = \overline{B}(s) \cap \left(\bigcap_{\{k: s \in \|\eta_k\|\}} \|\beta_k\| \right)$$

In words, $\bigcap_{\{k: s \in \|\eta_k\|\}} \|\beta_k\|$ is the set of behaviors allowed by the contract at state s, namely those behaviors that satisfy the instructions specified in all the clauses that apply to state s. The set $B^g(s)$ is then the set of behaviors that are feasible and allowed by the contract at state s. Once the contract is signed, the agent has to choose his behavior in set $B^g(s)$.

To simplify the analysis, we assume that a contract is enforceable only if it specifies feasible (hence, non-contradictory) obligations for all states. This is by no means the only reasonable assumption. An alternative assumption would be that courts enforce the contract in all states for which the contract specifies feasible obligations, and enforce no obligations in all other states. Under this alternative assumption, it is possible that the optimal contract will stipulate infeasible obligations (i.e. will contain contradictory prescriptions) in some states, as this may potentially economize on writing costs.

We say that contract g is *feasible* if $B^g(s) \neq \emptyset$ for all $s \in S$. We denote the set of feasible contracts by F.

2.6. Equivalent Contracts and Efficiently Written Contracts

An important element of our logical construct is the definition of an efficiently written contract. For this, however, we first need appropriate notions of equivalence between contracts.

Definition 2. *Two contracts g and h are* behaviorally equivalent *if they specify the same constraints on behavior at each state, that is, $B^g(s) = B^h(s)$ for all $s \in S$. Two contracts g and h are* equivalent *if they are behaviorally equivalent and they have the same cost ($C(g) = C(h)$).*

The following is our notion of efficiency in writing a contract:

Definition 3. *A contract g is* efficiently written *if $C(g) \leq C(h)$ for every behaviorally equivalent contract h.*

Consider, in the babysitting example, the contract

$$g = (e_1 \wedge e_2) \to (a_1 \wedge a_2)) \wedge ((e_1 \wedge \neg e_2) \to a_1) \wedge ((\neg e_1 \wedge e_2) \to a_2).$$

If $r > 0$, contract g is *not* efficiently written. Consider the alternative contract

$$h = (e_1 \to a_1) \wedge (e_2 \to a_2).$$

Both contracts instruct the babysitter to feed the baby if he cries and to change diapers if he smells, but h has a lower writing cost: $C(g) = C(h) + 6r$. This also shows how contracts with non-mutually exclusive contingencies can save on writing costs.

2.7. The Game

We consider a game where the principal proposes a feasible contract (g, t) (job description and transfer) to the agent, who can either accept or reject. If the contract is signed, after observing the realized state s the agent is supposed to behave according to some $b \in B^g(s)$. Implicit in this timing is the assumption that it is not feasible for parties to negotiate or communicate after the state is realized and before the agent takes action.[12] This kind of setting is realistic in a variety of situations, for example when the agent must react quickly to contingencies, especially if the principal is not around at that time (think of the babysitting case).[13]

The principal and the agent are *risk neutral.* For every (s, b), the principal gets gross benefit $\pi(s, b)$ and pays $t + C(g)$. Note that we assume that only the principal – who has all the bargaining power – pays the writing costs.[14] In case of disagreement the principal pays $C(g)$ and the agent gets utility $\overline{U}$.

The agent's preferences are represented by the utility function

$$U(s, b, t) = t - \delta(s, b),$$

where $\delta(s, b)$ is the disutility of behavior b in state s. Given contract g and state s, we let

$$BR^g(s) = \arg \min_{b \in B^g(s)} \delta(s, b)$$

denote the set of best responses of the agent at state s.

We assume that the state and the agent's behavior are verifiable in court, but preferences and realized payoff levels are not. If preferences were verifiable, the first-best outcome could trivially be achieved by a contract of the form "The agent's behavior b must maximize the sum of the parties' utilities."[15] On the other hand, if realized payoff levels were verifiable, the first-best outcome could be achieved by offering the agent a transfer that increases one-for-one with the principal's realized payoff level. We also assume that the principal cannot "sell the activity" to the agent (i.e. the agent cannot be made the recipient of the gross payoff π); this would be essentially equivalent to specifying a contingent transfer as in the previous point.

[12] Note that this assumption rules out message-based mechanisms a' la Maskin-Tirole (1999).

[13] See Battigalli and Maggi (2000) for an extension of this analysis to a setting where spot contracting is feasible and parties interact repeatedly.

[14] This assumption is not entirely innocuous. Anderlini and Felli (1997) show that, if both parties must incur a transaction cost before the negotiation takes place, it is possible that in equilibrium no contract will be signed even though it would be efficient to do so.

[15] Sometimes we do observe general "first-best" clauses of this kind, for example when a contract requires an employee to "act in the company's best interest." This kind of clause makes sense if the company's payoff function can, at least imperfectly and at a cost, be verified in court. A more general model would allow for imperfect and costly verification of payoff functions. In this case, it is conceivable that an optimal contract might include both a general "first-best" clause as well as specific behavioral clauses.

The principal and agent's prior beliefs about the exogenous state are given by a common probability measure $\mu \in \Delta(S)$. (With a slight abuse of notation we also write sentences and formulae as an argument of μ, as in $\mu(\eta) = \mu(\{s : s(\eta) = 1\})$.)

We also assume that there are gains from trade gross of writing costs:

Assumption 1. (*Gains from trade.*) There is some behavioral function $\mathbf{b} : S \to \{0,1\}^{\Pi^a}$ such that

$$\mathbf{b}(s) \in \overline{B}(s), \text{ for all } s \in S,$$
$$\sum_{s \in S} \mu(s) [\pi(s, \mathbf{b}(s)) - \delta(s, \mathbf{b}(s))] > \overline{U}.$$

2.8. Optimal Contracts

We say a feasible contract (g, t) is a subgame perfect equilibrium contract if there is a subgame perfect equilibrium where the agent accepts (g, t).[16] For definiteness, we focus on subgame perfect equilibria whereby the agent breaks ties in favor of the principal.

Clearly, a subgame perfect equilibrium contract (g, t) must satisfy the agent's participation constraint with equality, that is, $t = \overline{U} + \sum_{s \in S} \mu(s) \left(\min_{b \in B^g(s)} \delta(s, b)\right)$. Therefore from now on we will omit the transfer t from the notation, and identify the contract simply with its job description.

A contract $g^1$is a *first best* contract if it solves

$$\max_{g \in F} \left\{ \sum_{s \in S} \mu(s) \left[\max_{b \in BR^g(s)} \pi(s, b) - \min_{b \in B^g(s)} \delta(s, b) \right] \right\},$$

A contract g^2 is an *optimal* contract if it solves

$$\max_{g \in F} \left\{ \sum_{s \in S} \mu(s) \left[\max_{b \in BR^g(s)} \pi(s, b) - \min_{b \in B^g(s)} \delta(s, b) \right] - C(g) \right\}$$

Our assumptions and definitions have three immediate implications:

[16] Note that in general there may be a difference between the no-contract situation (status quo) and a contract specifying only a zero transfer (which we call 'empty' contract), because an empty contract gives the agent the right to affect the principal's payoff with his unrestricted behavior. However, in the applications we will consider the empty contract is equivalent to the status quo.

Remark 1. *(i) There are at least one optimal contract and one efficiently written first best contract. (ii) By Assumption 1, optimal contracts are subgame perfect equilibrium contracts and first best contracts are subgame perfect equilibrium contracts of the game where writing costs are zero. (iii) Every optimal contract is necessarily an efficiently written contract.*

The proof of remark 1, along with those of all the following results, are in Appendix. The following intuitive proposition says that if writing costs are very small, an optimal contract must be an efficiently-written first best.

Remark 2. *There is some $\overline{c} > 0$ such that, if $0 < c < \overline{c}$, then every optimal contract is an efficiently-written first best.*

This result is similar to one derived by Anderlini and Felli (1999, Proposition 6), though in a different setting.

2.9. Incompleteness of Contracts

We can distinguish between two basic forms of incompleteness: (1) We say that a contract g is *rigid* if there are two distinct states $s, s' \in S$ such that $B^g(s) = B^g(s')$. (2) We say that a contract g exhibits *discretion at* s if $\#B^g(s) > 1$ ($\#X$ denotes the cardinality of set X).

In words, a contract exhibits discretion if the behavioral correspondence B^g is not single-valued, and is rigid if B^g is not one-to-one. Rigidity is a lack of sensitivity of the contractual obligations to the external state. Our notion of discretion includes as a special case a notion of incompleteness that is fairly common in the literature, namely that the contract is "silent" (it specifies no obligations) at a given state.

Rigidity and discretion are two ways of saving on writing costs. Omitting from the contract an elementary sentence e_n saves on the cost of describing contingencies, but makes the contract rigid. Omitting from the contract an elementary sentences a_n saves on the cost of describing behavior, but gives discretion to the agent. A key objective of the analysis will be to examine under what conditions the optimal contract displays one or the other form of incompleteness.

Note that, in our main application, any rigidity or discretion implies a divergence from the first-best outcome. In a more general model this need not be true, as a contract exhibiting rigidity or discretion might implement the first-best outcome.

3. Application: a "Match-the-State" Setting

In this section we impose more structure on the framework. This will enable us to characterize the optimal contract and derive interesting comparative-statics predictions.

We assume that there is a one-to-one correspondence between elementary events and elementary activities. The principal wants elementary activity a_n to be carried out if and only if elementary event e_n occurs. For example, in our baby-sitting situation, if it is sunny the babysitter should take the baby for a walk (and if it is not sunny she should keep him at home), if the baby cries she should feed him (and if he does not cry she should not feed him), and so on.

The principal gets incremental benefit π_n from "matching" e_n with a_n (or $\neg e_n$ with $\neg a_n$), while he gets zero incremental benefit if there is a "mismatch". π_n may depend on n but is independent of other contingencies and activities.

Let $\Pi^e = \{e_1, ..., e_N\}$, $\Pi^a = \{a_1, ..., a_N\}$. We let N also denote the index set $\{1, ..., N\}$.[17] The principal's gross benefit function is:

$$\pi(s, b) = A_N \sum_{n=1}^{N} \pi_n[s(e_n)b(a_n) + (1 - s(e_n))(1 - b(a_n))]. \tag{3.1}$$

where $\{\pi_n\}_{n=1}^{\infty}$ is a fixed sequence of distinct, positive real numbers and $A_N = \frac{1}{\sum_{n=1}^{N} \pi_n}$. The reason we normalize gross benefits by A_N is that in this way, when we perform comparative statics with respect to N, we compare situations with the same potential gross benefit. We adopt the convention that elementary events and activities are ordered in terms of decreasing weight, that is,

$$\pi_n > \pi_{n+1} > 0, \ \forall n \geq 1. \tag{3.2}$$

[17] Note that we could equivalently assume $\Pi^a = \{a_1, \overline{a}_1, ..a_n, \overline{a}_n.., a_N, \overline{a}_N\}$ and $\bar{B}(s) = \{b : b(a_n) = 1 \text{ if and only if } b(\overline{a}_n) = 0, \forall n \in N\}$

The agent's interests are always in conflict with the principal's, in the sense that her preferred actions are always opposite the principal's preferred actions. This assumption is convenient because it implies that any first best contract is neither rigid nor loose, as we will see shortly. Therefore, any rigidity or looseness that arises in a contract is "excessive" (it implies an inefficiency). The agent's utility function is:

$$U(s,b,t) = t - \delta\pi(s,b),\ (0 < \delta < 1),\ \overline{U} = 0. \tag{3.3}$$

The parameter δ captures the strength of the conflict of interests between principal and agent; $(1-\delta)$ is a measure of the potential gain from contracting, or in other words the efficiency loss from noncooperative behavior.

The elementary events e_n, $n = 1, ..., N$, are i.i.d. with marginal probability p. By convention we assume $p > 1/2$ (we do not consider the knife-edge case $p = 1/2$ to avoid ties that would make the analysis more tedious). Formally we are assuming:

$$\mu\left((\bigwedge_{m\in K} e_m) \wedge (\bigwedge_{n\in N\setminus K} \neg e_n) \right) = p^{\#K}(1-p)^{N-\#K},\quad (p > \frac{1}{2}),\quad \forall K \subseteq N. \tag{3.4}$$

We can think of p as capturing the degree of uncertainty in the environment: the higher p, the lower the uncertainty (notice that the variance of the random variable $s(e_n)$ is decreasing in p).

We also assume that the parameters satisfy a genericity condition, to avoid ties:

$$\forall n \in N,\ \frac{1}{pA_N}\frac{c}{1-\delta} \neq \pi_n \neq \frac{1}{(1-p)A_N}\frac{c+2r}{1-\delta}. \tag{3.5}$$

In this model, the efficiently-written first best contract takes a very simple form:

Remark 3. *Under (3.1), (3.2) and (3.3), every efficiently-written first best contract is equivalent to:*

$$\bigwedge_{k=1}^{N} \left((e_k \to a_k) \wedge (\neg e_k \to \neg a_k) \right).$$

Notice that the efficiently-written first best contract is neither rigid nor loose. We will first characterize the structure of the optimal contract, then examine how it is affected by changes in the key parameters.

Proposition 1. *Under (3.1), (3.2), (3.3), (3.4) and (3.5), every optimal contract is equivalent to:*

$$\left(\bigwedge_{k=1}^{K_1}((e_k \to a_k) \wedge (\neg e_k \to \neg a_k))\right) \wedge \left(\bigwedge_{k=K_1+1}^{K_1+K_2}(\top \to a_k)\right)$$

where $K_1 = \max\left\{k \in N : \pi_k > \frac{1}{(1-p)A_N}\frac{c+2r}{1-\delta}\right\}$ *and* $K_1 + K_2 = \max\left\{k \in N : \pi_k > \frac{1}{pA_N}\frac{c}{1-\delta}\right\}$ *(by convention we let* $\max\emptyset = 0$ *and* $\bigwedge_{k=m}^{\ell}(\cdot) = \top$ *if* $m > \ell$*).*

The above proposition states that, in general, the optimal contract is characterized both by rigidity and by discretion. In particular, the set of N actions is partitioned in three groups: a group of more important actions (i.e. those associated with higher π_n) is regulated by contingent (double-)clauses of the form $(e_n \to a_n) \wedge (\neg e_n \to \neg a_n)$; a group of less important actions is regulated by rigid clauses of the form $(\top \to a_n)$; and the least important actions are left to the agent's discretion (i.e. they are not regulated at all). Any of these three subsets of actions may be empty, depending on parameters.

Next we sketch the basic intuition for the result. Given that payoffs are separable with respect to the dimensions (or aspects) $n = 1, ..., N$, and that elementary events are independent, we can focus on contracts with a "separable" structure, in the sense that each dimension n is handled by a clause that depends only on e_n and/or a_n. (It must be noticed, however, that this depends crucially on the assumption of conflict of interest; in section 4.4 we will see that, when the agent's interests are partially aligned with those of the principal, it is no longer true that the optimal contract can be derived by looking separately at each dimension n.) This means that we only need to choose the optimal clause for each n. There is only a small number of candidate clauses for each n. Look at the following table:[18]

k^{th}**Clause**	**Label**	**Incremental Net Surplus**
$e_k \to a_k$	$\tilde{C}_k$	$p(1-\delta)A_N\pi_k - 2c$
$\neg e_k \to \neg a_k$	$\tilde{C}_k^{\neg}$	$(1-p)(1-\delta)A_N\pi_k - 2c$
$(e_k \to a_k) \wedge (\neg e_k \to \neg a_k)$	C_k	$(1-\delta)A_N\pi_k - 2(c+r)$
$\top \to a_k$	R_k	$p(1-\delta)A_N\pi_k - c$
$\top \to \neg a_k$	$R_k^{\neg}$	$(1-p)(1-\delta)A_N\pi_k - c$
$\top \to \top$	D	0

Table 1

We have attached labels to clauses to simplify the notation. Label C stands for "contingent," R stands for "rigid", and D stands for "discretion." Any other clause about aspect k is clearly suboptimal, because it prescribes the wrong action. Next we need to select the clause with the largest incremental net surplus for each aspect k. By inspection of the table, one can verify that the simple contingent clauses $\tilde{C}_k$ and $\tilde{C}_k^{\neg}$ and the rigid clause $R_k^{\neg}$ cannot be optimal. The choice is thus narrowed down to clauses C_k, R_k and D, which cost respectively $2(c+r)$, c and zero.

Having narrowed down the choice in this way, the result that the most important tasks are regulated by contingent clauses and the least important tasks are left to the agent's discretion is very intuitive. The reason a task of intermediate importance is regulated by a rigid clause, on the other hand, is that a rigid clause "gets it right" with probability $p > 1/2$. The task is important enough that the expected benefit of a rigid clause outweighs the cost c, but is not so important that the benefit of a contingent clause (which "gets it right" with probability one) exceeds $2(c+r)$.

Using Proposition 1, it is easy to examine how changes in the key parameters affect the optimal contract. To simplify the presentation of the comparative-statics results, we assume here that the recalling cost is proportional to c: $r = r'c$, with $0 \leq r' \leq 1$. Also, let $y = c/(1-\delta)$ denote the writing cost relative to the potential gross surplus. The next corollary summarizes how changes in y, p and N affect the degrees of rigidity and discretion in the optimal contract. The degree of rigidity is captured by the number of rigid clauses, while the amount of discretion

[18] Note, we use the index n to denote a dimension of the problem, and the index k to denote a clause of the contract. In this setting, of course, there is a one-to-one correspondence between dimensions and clauses.

is captured by the number of missing clauses. The corollary also looks at the impact of N on the length of the contract.

Corollary 1. *Suppose that (3.1), (3.2), (3.3), (3.4) and (3.5) hold. Then*
(i) As y increases, the number of rigid clauses initially increases and eventually decreases. The number of missing clauses is nondecreasing in y. If y^R (resp. y^D) is the minimum level of y for which there are rigid (resp. missing) clauses in the optimal contract, then $y^R < y^D$. The same statements are true if y is replaced by N.
(ii) The number of rigid clauses is nondecreasing in p, and the number of missing clauses is nonincreasing in p.
(iii) Suppose $y < p\frac{\pi_2}{\pi_1+\pi_2}$ (or equivalently, that for $N = 2$ the optimal contract contains two clauses). Then the total number of clauses in the optimal contract is non-monotonic in N, reaching a peak for some intermediate level of N.

The optimal contract can respond in two ways to increasing complexity costs (in the sense of higher writing costs, c, and/or a higher dimensionality of the problem, N): it can leave more discretion to the agent (i.e. clauses are dropped from the contract), or it can be made more rigid (i.e. clauses are made noncontingent). According to Corollary 1(i), contracts tend to be more rigid for intermediate levels of complexity costs, and to leave more discretion for higher levels of complexity costs. In particular, for moderate levels of complexity costs the contract displays only rigidity, for higher levels of complexity costs it is characterized by both rigidity and discretion, and for very high levels of complexity costs the empty contract is optimal.

[Insert Figure 1 here]

Figure 1 illustrates the structure of the optimal contract as a function of y and N, for given $p > 1/2$ and r'. The (y, N) space is divided into six regions. In region FB, the first best contract is optimal; in region NC the empty contract is optimal. In the remaining regions, a letter C (resp. R and D) means that there are some contingent (resp. rigid and missing) clauses in the optimal contract. Thus, in regions R and C+R, the optimal contract displays rigidity but leaves no discretion; in regions R+D and C+R+D, the contract is characterized by rigidity as well as discretion.

Note that an increase in potential gross surplus from the relationship σ has identical consequences as a decrease in the writing cost, since these two parameters enter the problem only through the ratio $\frac{c}{1-\delta} = \frac{c}{\sigma}$.

We believe this prediction is potentially testable. At a broad level, we can think of two ways of going about this. First, one could take a cross-section approach, by looking at the variation of contracts within an industry. For example, our model predicts that when the value of trade (σ) is relatively small, the contract should be short and contain only a few rigid clauses, leaving substantial discretion to the parties; when the value of trade is higher, on the other hand, the contract should be longer and contain both rigid and contingent clauses, leaving less discretion to the parties.[19]

Another possibility would be to observe how contracts change over time in situations where the value of contracting (captured by σ) increases. For example, this might be the case in a growing industry, to the extent that the size of individual firms (not only the number of firms) tends to grow. One could then check whether the evolution of contracts is consistent with the model's predictions about increases in σ.

Part (ii) of the corollary focuses on the impact of uncertainty. When uncertainty decreases (i.e. when p increases), the amount of rigidity in the optimal contract increases, and the amount of discretion falls. This is intuitive: when uncertainty is lower the efficiency cost of ignoring low-probability events and writing rigid clauses is lower, hence the number of rigid clauses tends to be higher. Moreover, when uncertainty is lower, leaving discretion looks relatively less

[19] We took a casual look at the area of construction contracts, and what we saw seems consistent with our model. Hauf (1968) and Douglas and O'Neill (1994), for example, report the most frequent types of construction contracts used in the United States. Projects of smaller value are generally handled by contracts that are fairly simple and short. These short contracts typically contain only a limited set of noncontingent instructions, including a plan of the facility and a specification of the materials to be used, and leave much discretion to the constructor. On the other hand, bigger projects tend to be handled by longer contracts. These longer contracts give much less discretion to the constructor: they contain a fair number of noncontingent instructions, as well as several contingent clauses, describing for example what the contractor is supposed to do if the site conditions change, or describing the contingencies under which the owner can request a change in the construction specifications.

attractive than imposing a rigid rule, and no more attractive than writing a contingent rule, thus the number of missing clauses tends to be lower.

Note that, as uncertainty increases, the optimal contract contains fewer clauses, and it may even be simpler, in the sense of having a lower total complexity cost $C(g)$ (it is easy to show examples where this occurs). This should be contrasted with the prediction of more traditional transaction-cost models, such as Dye (1985). In these models, an increase in uncertainty typically leads to more complex contracts, because contract incompleteness can only take the form of rigidity. What makes a difference in our model is the interplay of rigidity and discretion, which is absent from models *à la* Dye.

Part (iii) of the corollary highlights another interesting prediction of the model: the length of the contract is maximal for intermediate levels of complexity (intermediate N). This seems consistent with casual empirical observations. Contracts regulating very simple activities are of course short, but so too are those that regulate very complex activities (such as, for example, the job of a university professor).

4. Robustness

In the previous section we made a number of assumptions on the payoff structure, on the form of contracts and on the language used to write contracts. In this section we discuss the robustness of our qualitative results to changes in these assumptions. We begin by relaxing the symmetry assumptions and the one-to-one correspondence between elementary actions and events. Next we discuss the implications of correlated elementary events, complementary tasks, and partial alignment of interests. We then consider the possibility of contingent transfers. Finally, we discuss how results are likely to change in the presence of more general languages.

4.1. More general payoffs with conflict of interests

Here we remove all the symmetry assumptions and the one-to-one correspondence between elementary actions and events. We retain only a minimum of assumptions to ensure that the problem is separable in the N tasks, in the sense that we can optimize the contract task by task. This requires that expected payoffs are separable in the N tasks *and* that there is full conflict of interests between the parties. Characterizing the optimal contract when the problem is not separable across tasks is a very difficult problem, which we do not know how to solve.

We can drop the assumption that the number of elementary actions equals that of elementary events, and we can replace the gross benefit function of the basic model with the following

$$\pi(b,s) = \sum_{n=1}^{N} \pi_n g_n(b_n; \mathbf{s}_n)$$

where b_n stands for $b(a_n)$, $\mathbf{s}_n \equiv (s_{n_1}, s_{n_2}, ...,) \equiv (s(e_{n_1}), s(e_{n_2}), ...,)$ is the set of elementary random variables that are relevant for task a_n, and $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N$ do not overlap (i.e., each elementary event is relevant for at most one task). We can also replace the assumption of i.i.d. elementary events with the weaker condition that the vectors $(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N)$ are mutually independent. For example, we could have $\pi(b,s) = \pi_1 g_1(b_1; s_2, s_3) + \pi_2 g_2(b_2; s_1, s_4, s_6) + \pi_2 g_2(b_2; s_5)$, where the vectors (s_2, s_3), (s_1, s_4, s_6) and (s_5) are mutually independent. The task-specific scaling parameter π_n captures the 'importance' of task n: an increase in π_n (holding all else equal) blows up the gain from contracting on task n. To avoid tedious ties, we assume that $\arg\max_b g_n(b_n, \mathbf{s}_n)$ is unique for all n.

The agent's payoff is still given by $U = t - \delta\pi(b,s)$, and the reservation utility is still zero for both players.

We refer to this setting as the *generalized match-the-state* model. The reason we did not conduct the whole analysis within this more general model is two-fold. First, as we will see shortly, the characterization of the optimal contract and the comparative-statics results are not as crisp as in the simple match-the-state model. Second, in the simpler version of the model we could capture the degree of uncertainty with a single parameter (p), whereas in this more general setting there is no simple way to examine the comparative-statics effects of changes in the degree of uncertainty.

Let us focus on task n. Performing this task ($b_n = 1$) is efficient for a certain subset of states, say $E_n^* \subseteq S$. An efficiently-written complete contract will then involve a set of clauses of the form $(\eta_n^* \to a_n) \wedge (\neg\eta_n^* \to \neg a_n)$, where η_n^* is an efficiently-written formula with truth set E_n^*. Note that the efficiently-written complete contract may be partially rigid.

Next we examine the optimal contract.[20] Without loss of generality, we label elementary actions in such a way that the rigid clause $(\top \to a_n)$ is preferred to the rigid clause $(\top \to \neg a_n)$ for each n. We also assume that this preference is strict for each n, to avoid knife-edge cases. In the optimal contract, each task is regulated by a clause of one of three types: (1) A contingent clause of the form $(\eta_n \to a_n) \wedge (\neg\eta_n \to \neg a_n)$. If c is sufficiently low, η_n will coincide with η_n^*, and the clause will implement the first best for task n; if c is higher, η_n may be a simpler formula than η_n^*, and the clause may not implement the first best. Note that, if η_n is a simpler formula than η_n^*, the clause *may*, but need not be, more rigid than the first best clause. Note also that a contingent clause costs *at least* $2c$. (2) A fully rigid clause $(\top \to a_n)$, which costs c. In what follows we refer to this clause simply as *the* rigid clause. (3) A discretionary (empty) clause, which is costless.

Consider first the robustness of Proposition 1. In this more general setting, tasks may not only differ in 'importance' (π_n), but also in a number of other ways, since we allow the function g_n to vary by task. For this reason, we cannot hope the result of Proposition 1 to hold exactly as stated. But the result still holds in a *ceteris paribus* sense:

Remark 4. *Consider the generalized match-the-state model. As π_n increases, holding everything else constant, the optimal clause for task n switches from discretionary, to rigid, to contingent.*

Proof: see Appendix.

Broadly speaking, then, we still have the result that tasks of high importance tend to be regulated by contingent clauses, tasks of intermediate importance tend to be regulated by rigid clauses, and the least important tasks tend to be discretionary.

As for the impact of changes in c and A on the efficient contract, the result of Proposition 1(i) continues to hold:

[20] To keep the exposition simple, we continue to refer to the optimal contract, i.e. the contract that maximizes the net surplus. In this extended setting, the optimal contract need not be a subgame perfect equilibrium, because it may yield a negative net surplus (in this case, in equilibrium no contract is signed). At any rate, the comparative-statics results we state for the optimal contract are valid also for any contract signed in a subgame perfect equilibrium, in the relevant parameter region.

Remark 5. *In the generalized match-the-state model, Proposition 1(i) holds as stated.*

Proof: see Appendix.

Intuitively, the reason our comparative-statics results are robust is that the rankings between a contingent clause, a rigid clause and a discretionary clause in terms of writing costs and in terms of expected benefits have not changed: a contingent clause costs at least $2c$, a rigid clause costs c, and a discretionary clause costs nothing; on the other hand, the rankning in terms of expected benefits is reversed (in particular, the assumption of conflict of interests implies that a rigid clause yields a higher expected benefit than discretion).

Proposition 1(ii) cannot easily be extended to the generalized match-the-state model, because the impact of uncertainty can no longer be gauged by a single parameter. However, we believe that the main insight – that redcing uncertainty tends to increase rigidity – should still hold.

An important question is, to what extent our results hold when the parties' interests are partially aligned, when elementary events are correlated, or when payoffs are not separable across tasks. The following subsections characterize the otpimal contracts in simple examples with these features, showing that, in general, we cannot optimize the contract task by task. We do not have techniques to solve the general optimization problem in the non-separable case. But we conjecture that our qualitative comparative-statics results would survive, at least as tendencies. This conjecture is motivated by the following intuition. Introducing rigidity in the contract saves on the cost of describing contingencies, while introducing discretion saves on the cost of describing contingencies *and* on the cost of describing actions. This insight is quite general, and is the main driving force of our comparative-statics results.

4.2. Correlated events

If the assumption of independent elementary events (3.4) is dropped, so that elementary events are allowed to be correlated, the optimal contract is more likely to be rigid. Consider the special case $N = 2$, and let the probability distribution over the possible states be represented by the following table

	e_1	$\neg e_1$
e_2	$\frac{1}{2}-\varepsilon$	ε
$\neg e_2$	ε	$\frac{1}{2}-\varepsilon$

where $0 < \varepsilon < 1/2$. The entries of the table are the joint probabilities of the four possible states. Here, if $\varepsilon \neq 1/4$, events e_1 and e_2 are correlated (positively if $\varepsilon < 1/4$, negatively if $\varepsilon > 1/4$). The remaining assumptions of the model are unchanged.

Consider for example the case of perfect positive correlation ($\varepsilon = 0$). In this case, if c is sufficiently small, the contract $(e_1 \to a_1 \wedge a_2) \wedge (\neg e_1 \to \neg a_1 \wedge \neg a_2)$ is optimal, because it implements the first best outcome (the same is true for the contract $(e_2 \to a_1 \wedge a_2) \wedge (\neg e_2 \to \neg a_1 \wedge \neg a_2)$). These contracts are rigid, since they do not discriminate with respect to one of the two elementary events. Intuitively, if elementary events are perfectly correlated, there is effectively a single relevant elementary event, hence there is no need to discriminate with respect to both. If the event "the baby cries" is perfectly correlated with the event "the baby smells", you might as well give instructions based only on whether the baby cries. More generally, one can show that, as the degree of correlation increases (i.e. ε moves toward either extreme), the parameter region where the optimal contract is rigid becomes larger.

This observation is consistent with the result of Corollary 1(ii), that the optimal contract is more likely to be rigid if uncertainty is lower (in this case, if $|\varepsilon - \frac{1}{4}|$ is higher). The general point is that rigid contracts tend to be a more attractive means of saving on writing costs in the presence of *low-probability states*, whether these are due to asymmetric marginal probabilities or correlation across elementary events.

4.3. Complementary tasks

We have so far assumed that payoffs are separable across tasks. An interesting question is whether the presence of complementarities among tasks, which many argue to be a pervasive feature of modern manufacturing jobs,[21] makes contracts more or less incomplete.

Intuition might suggest that complementarities make the optimal contract more incomplete,

[21] See for example Milgrom and Roberts (1990).

because if some tasks are specified in the contract, the agent may "spontaneously" perform other tasks that are complementary to the ones pinned down in the contract. This intuition turns out not to be correct. We will present a simple example in which complementarities make the optimal contract *less* incomplete, provided c is relatively low and the marginal effort cost of each task is positive.

Suppose there is a single relevant state, so that the contract only needs to specify the agent's behavior, and only two elementary tasks. Define elementary tasks so that the first-best behavior is given by $b(a_k) = 1$, $k = 1, 2$. To understand the role of complementarities, it is useful to distinguish between complementarities in the revenue function and complementarities in the agent's utility. For this distinction to be meaningful, we allow for a more general utility function than in section 3, namely $U = t - \delta(b)$, where $b = (b(a_1), b(a_2))$. Let $\pi(b)$ denote the revenue function.

We start by setting up the benchmark case of no complementarities, i.e. of separable revenue and utility functions. Our separable revenue function is given by $\pi(b) = \pi_1 b(a_1) + \pi_2 b(a_2)$, where $\pi_1 \geq \pi_2$. The separable (dis)utility function is given by $\delta(b) = \delta b(a_1) + \delta b(a_2)$, where $\delta > 0$.

Coming to the case of complementarities in the revenue function, we assume

$$\pi(b) = g(\pi_1 b(a_1), \pi_2 b(a_2))$$

where $g(\cdot)$ is symmetric, increasing and supermodular (note that this implies $\pi(b)$ is also supermodular). To facilitate the comparison with the no-complementarity case, we also assume that $g(\cdot)$ is homogenous of degree one and that the potential revenue is the same in the two cases, $g(\pi_1, \pi_2) = \pi_1 + \pi_2$.

Clearly, the only candidates as optimal contracts are, in order of increasing incompleteness (looseness): (1) $(a_1 \wedge a_2)$, (2) (a_1), and (3) the empty contract. Also note that the agent will perform only the tasks (if any) specified in the contract. The following result obtains:

Proposition 2. *There exists a critical level of the writing cost c^* (function of the other parameters) such that, if $c < c^*$, the optimal contract when $\pi(b)$ is supermodular is (weakly) less incomplete than when $\pi(b)$ is additively separable. The opposite is true if $c > c^*$.*

The intuition for this result is clearest if tasks are strongly complementary, so that it is optimal to treat them as a group. In this case, the optimal contract has a bang-bang structure: if c is lower (higher) than a critical level, both tasks (no tasks) are specified in the contract. At this critical level of c, in the no-complementarity case the contract specifies only task a_1. It follows that task complementarities make the contract less incomplete if c low, and more incomplete if c is high. If complementarities are slight, the optimal contract may not have a bang-bang structure, in the sense that there is an interval of c for which the optimal contract specifies only task a_1; but it turns out that the result holds also in this case.

Next we discuss the case in which tasks are complementary in the agent's utility function. This may be the case if there are economies of scope, in the sense that performing one task reduces the effort required to perform the other task. The first remark is that supermodularity of $U(b)$ (or submodularity of $\delta(b)$) translates into supermodularity of the joint surplus function, i.e. $\pi(b) - \delta(b)$. For this reason, the implications of effort complementarities are identical to those of revenue complementarities, *unless* they are so strong that the marginal effort cost of a task is *negative*. For example, this is the case for task a_2 if $\delta(1,1) < \delta(1,0)$. Then, if task a_1 is specified in the contract, the agent will perform task a_2 even if it is not required by the contract. In this case, the optimal contract will be loose for any $c > 0$, even if c is very small. Thus, when effort complementarities are so strong that the marginal effort cost of some task is negative, the contract is looser than in the separable benchmark for low c; for higher c, the comparison is ambiguous.

4.4. Partially aligned interests: a "hold-up" example

In this subsection we present an example where interests are partially aligned. This is a version of our model that we can interpret as a hold-up model a' la Grossman-Hart-Moore [see Hart (1995, Ch. 4) and references therein].

Suppose that the agent can make relationship-specific investments whose optimal specification depends on the external state. There are N elementary investment activities a_k. Each a_k costs δ to the agent and yields incremental gross surplus π_k if event e_k occurs and zero otherwise. Thus, the first-best requires undertaking investment a_k if and only if event e_k occurs, provided $\pi_k > \delta$. For example, efficiency may require the agent to learn German if and only if Germany becomes a strategic market for the firm. Ex post, the parties bargain over the returns to the investments. They foresee that the bargaining will lead to a split of the returns where the agent captures a share equal to θ. Since the agent appropriates only a fraction of the returns to his investment, in the absence of contracts he would tend to "underinvest".[22] We allow parties to contract on the investment actions to mitigate or eliminate the hold-up problem.

Again, the difference between this setting and the one of section 3 is that it is no longer true that we can optimize separately for each dimension n. Now it may be optimal to have clauses of the form $(\top \to a_k \vee a_j)$. The reason is that now the interests of the parties are partially aligned, so it may be optimal to constrain the agent to execute *some* activities and give him some discretion regarding which ones to carry out, because this may induce him to execute the right activity under the right contingency. For this reason, solving the general $N-$dimension case is fairly complicated. However, the $N = 2$ example will be sufficient to make our basic point.

The payoff functions of the principal and the agent are respectively

$$\pi(s,b) = (1-\theta)\sum_{n=1}^{2}\pi_n s(e_n)b(a_n), \text{ with } \pi_1 \geq \pi_2,\ 0 < \theta < 1. \tag{4.1}$$

and

$$U(s,b,t) = t + \theta\sum_{n=1}^{2}\pi_n s(e_n)b(a_n) - \sum_{n=1}^{2}\delta b(a_n)\ ,\ (\delta > 0),\ \overline{U} = 0. \tag{4.2}$$

Assume that $\mu(.)$ is uniform, i.e.,

$$\mu(e_n) = \frac{1}{2},\ n = 1, 2 \tag{4.3}$$

As in section 3, $\bigwedge_{k=1}^{2}(e_k \to a_k) \wedge (\neg e_k \to \neg a_k)$ is a first best contract. Note however that, in this specification, when $\neg e_k$ occurs the agent does not want to execute a_k, because it is costly.

[22] We are implicitly assuming that the realized returns to investments are not verifiable. This is consistent with the assumption, made in the general framework, that realized payoff levels are not verifiable. If they were, the hold-up problem could be trivially solved by a contingent transfer that increases one-for-one with the investment returns. Recall also the assumption that the principal cannot sell the firm to the agent. This would be equivalent to a contingent transfer of the kind just described.

Therefore the first best contract with minimum writing costs is $\bigwedge_{k=1}^{2}(e_k \to a_k)$.

We will characterize the optimal contract for the symmetric case $\pi_1 = \pi_2 = \pi$; we will later remark on how asymmetries affect the results. Assume $\pi > \delta$, so that it is efficient to execute both investment activities, and $\pi < \delta/\theta$, so that there is a hold-up problem (meaning that, in the absence of contract, the agent would not execute either investment activity). Letting $\tilde{\delta} \equiv \delta/\pi$, we are thus assuming

$$\pi_1 = \pi_2 = \pi \text{ and } \theta < \tilde{\delta} < 1 \tag{4.4}$$

It is not hard to show that only the following contracts can be optimal

Contract	**Label**	**Net Surplus**
$(e_1 \to a_1) \wedge (e_2 \to a_2)$	FB	$\pi - \delta - 4c$
$\top \to a_1 \wedge a_2$	RR	$\pi - 2\delta - 2c$
$\top \to a_1 \vee a_2$	R/D	$\frac{3}{4}\pi - \delta - 2c$
$\top \to \top$	D	0

Table 2

Label "RR" indicates a fully rigid contract, and "R/D" indicates a contract that displays rigidity and leaves some discretion as well. Note that the contracts $(e_1 \to a_1)$ and $(\top \to a_1)$ cannot be optimal with $\mu = 1/2$ and $\pi_1 = \pi_2$, but they can be optimal in the asymmetric case, as we will discuss shortly. Also note that the contract $(e_1 \to a_1 \vee a_2)$ cannot be optimal, as it is always dominated by the simple contingent contract $(e_1 \to a_1)$.

It is immediate to see that, if there is rigidity in the optimal contract, the agent will *overinvest* in some states. Moreover, if the optimal contract is RR, the agent will overinvest (weakly) in *all* states. Having narrowed down the set of candidate contracts, it is a small step to characterize the structure of the optimal contract. Let $\tilde{c} \equiv c/\pi$ denote the writing cost relative to the gross surplus.

Remark 6. *Suppose that (4.1), (4.2), (4.3) and (4.4) hold.*
(i) If $\theta < \tilde{\delta} < \frac{1}{4}$, as $\tilde{c}$ increases the optimal contract goes from FB, to RR, to D.
If $\max\{\theta, \frac{1}{4}\} < \tilde{\delta} < \frac{1}{2}$, as $\tilde{c}$ increases the optimal contract goes from FB, to R/D, to D.

If $\max\{\theta, \frac{1}{2}\} < \widetilde{\delta} < 1$*, as* $\tilde{c}$ *increases the optimal contract goes from* FB *to* D*.*
(ii) If the empty contract is optimal, the agent underinvests (weakly) in all states. If contract R/D *is optimal, the agent underinvests in some states and overinvest in others. If* RR *is optimal, the agent overinvests (weakly) in all states.*

Our assumptions of uniform μ (maximal uncertainty) and $\pi_1 = \pi_2$ make for sharp results, but the qualitative insights remain valid in the asymmetric case as well. If μ is not uniform, the tendency to have rigidity in the contract is strengthened, since the presence of low-probability events reduces the cost of rigid rules. If $\pi_1 > \pi_2$, on the other hand, the main change in results is that there is a parameter region where the simple contingent contract ($e_1 \rightarrow a_1$) is optimal, and a parameter region where the simple rigid contract ($\top \rightarrow a_1$) is optimal.

Also note the role of the profit-sharing parameter θ. Recall that we assumed $\theta < \tilde{\delta} < 1$, for the problem to be interesting. Therefore, if the agent's share θ is lower than 1/2, there is a parameter region where the optimal contract is rigid, hence the agent overinvests; if the agent's share is higher than 1/2, the optimal contract is either the first-best contract or the empty contract. Thus, the possibility of overinvestment arises only when the agent's share of the surplus is small.

This result is interesting if contrasted with the approach – fairly standard in the hold-up literature – of simply assuming away contracts on investments. In the no-contract case, the prediction of the model is that the agent underinvests. One might have conjectured that, when writing costs are intermediate, so that it is optimal to write a partially incomplete contract, the agent will still underinvest, although to a lesser extent than in the no-contract case. As our result shows, this intuition is not correct. We have seen that, if the agent's share of the surplus is less than 1/2, it may be optimal to write a rigid contract, in which case the agent will overinvest in some states, and for some parameter values in *all* states.[23]

[23]We believe that our qualitative results do not depend on the zero-one nature of the investments. Suppose the relevant investment is to learn German, and this can be done at two levels of proficiency, a_1 and a_2, with respective effort costs of δ_1 and δ_2 (level 2 being the more advanced one). Describing the required proficiency level is costly; this is captured by the writing cost c. Suppose e_1 is the event "Germany becomes an important market for the firm," and e_2 is the event "The only other German-speaking employee quits." If both these events occur, it is efficient for the agent to do (only) a_2; if only one of the two events occur, it is efficient to do (only) a_1, and if neither event occurs it is efficient to do nothing. In this example, one can find a payoff

4.5. More general contract form

Thus far we have restricted our attention to contracts that specify a fixed transfer and a set of clauses of the form $\bigwedge_{k=1}^{K}(\eta_k \rightarrow \beta_k)$. We can call this a *forcing contract.* A more general contract would allow transfers to be contingent on actions and states. In this section we argue that, given our assumptions of verifiable states and behavior, risk neutrality and conflict of interests, there is no loss of generality in ignoring contingent transfers.

Consider the generalized match-the-state setting with conflict of interests, where $\pi(b,s) = \sum_{n=1}^{N}\pi_n g_n(b_n;\mathbf{s}_n)$, with $\mathbf{s}_1, \mathbf{s}_2, .., \mathbf{s}_N$ disjoint and mutually independent sets of elementary random variables. We assume that, for each n, g_n is *generic*, i.e. $g_n(0;\mathbf{s}_n) \neq g_n(1;\mathbf{s}_n)$ for all $\mathbf{s}_n$.

An *incentive contract* is a set of clauses $\bigwedge_{k=1}^{K}(\varphi_k \rightarrow t_k)$, where $t_k \in \mathbf{R}$ and φ_k is a formula about the environment, behavior or both.[24] We continue to assume that the cost of writing a contract is proportional to the number of distinct elementary sentences occurring in its clauses, while writing transfers has *per se* no cost. Note that an incentive contract is more general than a forcing contract: a forcing clause expressing a certain instruction can always be written as a contingent transfer that specifies that the agent gets a stiff penalty if the instruction is not followed.

Proposition 3. *Suppose that there are no recalling costs ($r = 0$). Then for any incentive contract, there exists a forcing contract that is not more costly to write and yields at least the same surplus in every state.*

Proof: see Appendix.

The intuition for this result is the following. First note that, due to the assumptions of verifiable state and behavior and of risk neutrality, the only possible benefit of contingent transfers is to save on writing costs. Now suppose one wants to implement the efficient contingent rule

function $\pi(s,b)$ and parameters c, δ_1 and δ_2 such that the optimal contract is $(\top \rightarrow a_2)$, which entails (weak) overinvestment by the agent in all states.

[24] Our language can be enriched so as to include countably many elementary sentences of the form "the principal transfers t dollars to the agent" where t is a rational number.

for a given task. With a forcing contract, this requires describing both the task and the external events. In principle it might be possible to implement the same rule by specifying a transfer contingent on the agent's behavior but not on the external events; this would save on writing costs. However, due to the conflict of interests, it turns out that it is never possible to implement an efficient contingent rule without making the transfer directly contingent on the external events.

Of course we are not suggesting that contingent transfers are not important in reality. They are indeed used very frequently. Our point here is simply that there is no role for contingent transfers given the payoff structure structure of our basic model and assuming $r = 0$.

The assumption that there is no cost in "recalling" a primitive sentence already used in the contract also plays a role. One can construct examples where a contingent transfer saves on such "recalling costs," however this is no longer possible if there is a small cost of specifying transfers *per se*, or if the language includes the connective "$\leftrightarrow$" ("if and only if").

A role for contingent transfers may also appear if we consider a more general payoff structure. Consider a general "hold-up" model with N tasks, where interests are partially aligned:

$$\pi(s,b) = (1-\theta)\sum_{n=1}^{N}\pi_n s(e_n)b(a_n),\ U(s,b,m) = m + \theta\pi(s,b) - \delta\sum_{n=1}^{N} b(a_n)$$

In this case, for some parameters it is optimal to make transfers contingent on elementary actions. To see this, note that the following contract implements (strictly) the first-best outcome: $g = \bigwedge_{k=1}^{N}(a_k \rightarrow t_k)$ where $\delta - \theta\pi_k < t_k < \delta$. Indeed, this is the efficiently-written first-best contract, therefore it is optimal for c sufficiently small.

We conclude this subsection by proving rigorously a claim made in the introduction, namely that Dye's (1985) approach to modeling complexity costs cannot explain the presence of discretion.

As a preliminary remark, note that, in our basic model, a contract g induces a behavioral correspondence $B^g : S \rightarrow 2^B$ (where $B = \{0,1\}^{\Pi^a}$). If we apply Dye's approach to this correspondence, in order to compute the cost of writing g we simply have count the number of distinct values (subsets of B) attained by B^g. It does not matter whether these subsets are singletons or not. Therefore any contract g with discretion is (weakly) dominated by a contract h that exhibits no discretion (i.e. such that B^h is a function). In other words, Dye's approach cannot account for the cost of describing behavior if it is applied to functions (or correspondences) from S to B.

However, if we consider a more general incentive contract $g = \bigwedge_{k=1}^{K}(\varphi_k \rightarrow t_k)$, we see that it induces a function $T^g : S \times B \rightarrow \mathbf{R}$ that specifies for each pair (s,b) a monetary transfer $T(s,b) = \sum_{k:(s,b)\in\|\varphi_k\|} t_k$. Therefore we can apply Dye's hypothesis on writing (or complexity) costs to T^g. In this case, the optimal contract according to Dye can take only two forms: (i) if the cost of a two-valued function is not extremely high, then the optimal contract is equivalent to a forcing contract that specifies a high penalty $-P$ if b is not first best given s and a positive transfer t otherwise, (ii) in the extreme case where the cost of a two-valued function is very high the optimal contract is empty.

4.6. Alternative languages

A key aspect of our approach is that we model explicitly the language used to write contracts. A natural question then arises: How robust are the predictions of the theory to changes in the language, and in particular in the set of primitive sentences?

We postulated a set of primitive sentences that seems natural given the structure of payoffs. Payoffs depend on a set of binary random variables $(s_1, s_2, ...)$ and on a set of binary choice variables $(b_1, b_2, ...)$, so we have assumed that each of these variables is associated with a primitive sentence (in what follows we refer to this language as the "natural" language, and to its primitive sentences as "natural" primitive sentences). Still, it is important to think about the implications of alternative sets of primitive sentences.

We start with a remark that should put things in perspective. Our results cannot hold for *all* possible sets of primitive sentences. Consider an extremely rich language which associates a primitive sentence to each possible complete contract: for example, contract A, contract B, etc. We denote this language as $\mathcal{L}^{FB}$. With language $\mathcal{L}^{FB}$, the parties could always write a complete contract at the cost of c, and we would have no contract incompleteness (if c is not too large).

However, we believe a language like $\mathcal{L}^{FB}$ is highly unrealistic. The common-knowledge language of a society, in the sense of the language understood by the society's courts, must serve a large population of heterogenous contracting parties. Thus, language $\mathcal{L}^{FB}$ would have to include a primitive sentence for each *conceivable* complete contract. The number of conceivable complete contracts in reality is astronomical. If there is a social cost of having a richer language (because a richer language is more costly to learn, to teach, to remember, and to codify in vocabularies), then language $\mathcal{L}^{FB}$ will be excessively costly.

Another "salient" language that one might consider is the one that associates a primitive sentence to each possible contingency (i.e. subset of S) and to each possible set of behaviors (i.e. subset of B). Also this language, which we denote $\mathcal{L}^{W}$, is unreasonably rich. To get an idea of the dimensionality of languages $\mathcal{L}^{FB}$ and $\mathcal{L}^{W}$, and how they compare with the natural language, let us be more explicit about the heterogenous population of contracting pairs that we have in mind.

There are L potentially relevant binary random variables s_n ($s_n \in \{0,1\}$, $n = 1, ..., L$) with joint distribution $\mu \in \Delta\left(\{0,1\}^L\right)$ and K binary choice variables b_k ($b_k \in \{0,1\}$, $k = 1, ..., K$). There is a continuum of contracting pairs. Each contracting pair m is characterized by a surplus function $\sigma^m : \{0,1\}^L \times \{0,1\}^K \to \mathbf{R}$. Unlike the previous sections, here we impose no structure on payoffs. In fact, we assume that the surplus function σ^m is distributed with full support across the population; in other words, any surplus function is possible. We also assume $L = K$, to ease computations. A propositional language $\mathcal{L} = (\Pi, \|\cdot\|)$ assigns elementary sentences π_1, π_2, ... to subsets W_1, W_2, ... of $W = \{0,1\}^L \times \{0,1\}^L$ with the interpretation that W_j is the truth set of π_j, that is, $W_j = \|\pi_j\|$.[25]

Language $\mathcal{L}^{FB}$ assigns a primitive sentence to each possible first-best mapping. Since any surplus function σ^m can occur in the population, the set of possible first-best mappings is simply the set of all possible behavioral functions $b_s : \{0,1\}^L \to \{0,1\}^L$. Then it is not hard to show that language $\mathcal{L}^{FB}$ includes $\left(2^L\right)^{2^L}$ primitive sentences. The dimensionality of language $\mathcal{L}^{W}$ is lower but still mind-numbing, as it contains 2^{1+2^L} primitive sentences. In contrast, the natural language contains only $2L$ primitive sentences. Therefore, if the world is "big", in the sense that

[25] Formulas in $\mathcal{L}$ are obtained from elementary sentences using parentheses and logical connectives and their truth sets are obtained in the usual way (i.e., $\|\neg\varphi\| = W \backslash \|\varphi\|$, $\|\varphi \wedge \psi\| = \|\varphi\| \cap \|\psi\|$, etc.).

L is large, and the cost of a language is proportional to the number of its primitive sentences, the cost of $\mathcal{L}^W$ and $\mathcal{L}^{FB}$ overwhelms that of the natural language. Of course there are benefits associated with richer languages such as $\mathcal{L}^W$ and $\mathcal{L}^{FB}$, because they allow contracting parties to save on writing costs, but to the extent that such benefits are bounded as L increases, the natural language will dominate $\mathcal{L}^W$ and $\mathcal{L}^{FB}$.

To conclude this part of the discussion, we think the interesting question is not whether the theory's predictions are robust to an arbitrary choice of language, but whether they are robust to *reasonable* alternative languages.

To give the reader a sense of direction, we anticipate where we are heading. We are going to argue that a reasonable class of languages to focus on has the following two-tier structure: a basis made of natural language, plus a set of additional primitive sentences that denote the more frequently used formulas.

4.6.1. Two-tier languages

As a first step we argue that the natural language has a desirable efficiency property, which makes it a reasonable candidate for a general-purpose language, and a basis for further language enrichments. But first we need a premise.

Since we think of the size of the world (indexed by L) as very large, and we view the cost of a language as increasing in the number of its primitive sentences, then the language is likely to be incomplete, in the sense of not being able to describe the whole world. It makes sense, then, to talk about the *expressive power* of a language. Formally, we define the expressive power of a language as the number of subsets of W (we can call these *events*[26]) that the language can describe. This is a very broad notion of expressive power, that gives the same importance to all events. In the area of contracting, describing some events may be more important than describing others, but we have in mind that a language is used not only to write contracts, but to communicate a wide variety of messages in a wide variety of circumstances.

In what follows, at the risk of abusing our terminology, we will broaden the meaning of

[26] Note that we are slightly changing our terminology: "events" now concern both environment and behavior.

the expression "natural language": we will call "natural language" any collection of natural primitive sentences, even if this collection is not complete.

Remark 7. *Any natural language has the following property: there is no language with higher expressive power and the same number of primitive sentences, or with the same expressive power and fewer primitive sentences.*

Proof: See Appendix.

In what follows we refer to the property described in the remark as "minimality." Note that the two alternative languages we considered above, namely $\mathcal{L}^W$ and $\mathcal{L}^{FB}$, are not minimal. Another example of language that does not satisfy this property is the language that assigns a primitive sentence to each state and to each behavior.

Intuition for this result can be gained by considering a simple example. Suppose there is only one elementary event and one elementary action. Then the world W contains only four states, say $\omega_1, ..., \omega_4$. Suppose we need to choose a language that contains only two primitive sentences. Intuitively, the way to maximize expressive power is to choose the two primitive sentences so that (i) each splits the set W in half, and (ii) they are orthogonal (in the sense that, conditional on the truth value of one primitive sentence, the other primitive sentence splits the world in half). The natural language satisfies both of these properties. An example of language that does not satisfy them is the one that assigns the two primitive sentences to two states of the world (singletons).

We believe that the natural language, by virtue of its minimality, is a plausible candidate for a general-purpose language. In what follows, we will think of the natural language as the starting point from which further language enrichments can be developed. Short of a full theory of language evolution, which is beyond the scope of this paper, we think this is a plausible approach for the purposes of our discussion.

In reality, not all events need to be described with the same frequency. Suppose that a particular formula φ (which might describe for example an external contingency, or a contract clause) turns out to be used very frequently. Then it may be efficient to denote φ with a new primitive sentence; this will save on writing costs (or more generally on communication costs), and these savings may outweigh the social cost of increasing the size of the language.

For example, if e_1 : "humid weather" and e_2 : "hot weather", and the formula $e_1 \wedge e_2$ occurs frequently, this may be denoted by a new primitive sentence, $\tilde{e}$: "tropical weather." Or, if a_1 : "watch television with the baby" and a_2 : "sing to the baby," the formula $a_1 \vee a_2$ can be denoted by a new primitive sentence, $\tilde{a}$: "entertain the baby." Similarly, if a particular contract clause, say $\eta_j \longrightarrow \beta_k$, is frequently used, this can be denoted by a new label. Even a whole contract may be frequent enough to warrant the addition of a new primitive sentence in the vocabulary.

We can now come back to the question that motivated this section: how robust are the key predictions of our model to reasonable alternative languages? If language has the two-tier structure that we have argued for, then our comparative-statics results are valid subject to an important qualification: they apply more tightly the less standard is the contracting problem. Next we make this statement a bit more precise.

We believe that our results are robust to the relabeling of standard formulas about the environment (as in the example of $\tilde{e}$: "tropical weather" mentioned above) or about behavior (as in the example of $\tilde{a}$: "entertain the baby"). The intuition is the usual one: even with this relabeling, it is still true that rigidity saves on the cost of describing contingencies, while discretion saves both on the cost of describing contingencies and on the cost of describing actions.[27] Results may change, however, if the contracting problem features a substantial number of standard *clauses*, i.e. formulas of the form $\eta \rightarrow \beta$ that can be replaced by simple labels, because in this case contingent clauses may not be more costly than rigid clauses. Broadly speaking, if the parties can use standard clauses for some aspects of the contract, then our analysis can be applied only to the non-standard part of the contract.[28]

[27] We can prove this claim rigorously in a particular case. Consider a contracting problem as in section 4.1. Suppose that, in addition to the natural language, there is an additional set of primitive sentences $\tilde{e}_k$, $k = 1, .., n^e$, that can replace more complex formulas φ_k about the environment. Furthermore, suppose that every replaced formula φ_k involves elementary events that are relevant for only one task. Then the contracting problem can still be analyzed task by task and the results of section 4.1 hold as stated, because it is still true that a contingent clause costs at least $2c$ while a rigid clause costs only c. In a more general setting, the introduction of standard formulas about the environment or behavior may break the separability of the problem. This is the reason for our rather cautious claim in the text.

[28] There is a subtle but important distinction to make. We are talking about situations where parties can take advantage of *existing* standard contract clauses, not about the *creation* of a standard contract, in the sense for

We conclude this section with a remark on "private" languages. Consider two contracting parties that face a non-standard problem, so that they can only use the natural language to write an enforceable contract. Can they save on writing costs by creating new primitive sentences? Given our assumption that the cost of writing a contract is proportional to the number of *distinct* primitive sentences that appear in the contract, the answer is no. The reason is that, in order for the courts to understand the contract, any new primitive sentence needs to be defined *within the contract* in terms of the common-knowledge language, and doing so is at least as costly as writing the contract in the common-knowledge language. If we had a positive cost r of 'recalling' primitive sentences within the contract (see footnote **??**), then the creation of new primitive sentences could save on writing costs, but the comparative-statics results would be very similar to the ones we presented.

5. Unforeseen events

In this section we discuss how the model can be extended to allow for unforeseen events. We start with a preliminary consideration. There are two types of unforeseen events: unforeseen aspects of the *environment* and unforeseen aspects of *behavior*. Even though the latter notion is rarely emphasized in the literature, we would argue that it is quite relevant in contexts where the complexity of behavior is an important issue. Often, when two parties face a non-standard contracting situation, they have to think very hard about all the possible ways that each party can take advantage of the other, so that these actions can be prohibited by the contract. In what follows we will use the expression "unforeseen events" to encompass both aspects of the environment and of behavior.

The following is a simple way to extend our framework to allow for unforeseen events. Suppose that, in addition to the elementary exogenous events (binary random variables) that the parties have in mind, $s_1, ..., s_N$, there is an additional set of "latent" elementary events, $s'_1, ..., s'_{N'}$, that the parties do not have in mind because they are normally turned "off." Examples of latent elementary events might be the appearance of Internet (for someone living a

example of a company drafting a contract to be offered to multiple customers. Our analysis is broadly applicable to the latter type of situation, as long as the contract-writer faces a fresh contracting problem.

few decades ago). Ex post, the parties may become aware of these elementary events if they are turned "on." By convention, let us identify the "off" state of a latent elementary event s'_j with the value $s'_j = 0$. Similarly, we can suppose that there is a set of latent elementary actions (binary choice variables), $b'_1, ..., b'_{M'}$, that the principal does not have in mind when drafting the contract, because they are normally "off."[29] A latent elementary action might be "the agent gets an autotransfusion"[30]. We identify the "off" state of a latent elementary action b'_j with the value $b'_j = 0$.

If the "true" benefit function is $\tilde{\pi}(s_1, .., s_N, s'_1, .., s'_{N'}; b_1, .., b_M, b'_1, .., b'_{M'})$, we can think of the principal as having a "perceived" benefit function at the time of writing the contract, given by $\pi(s_1, .., s_N; b_1, .., b_M) = \tilde{\pi}(s_1, .., s_N, 0, .., 0; b_1, .., b_N, 0, .., 0)$. From the point of view of ex-ante perceived payoffs, the optimal contract will be the same as the one we characterized. From the point of view of the "true" payoffs, however, the presence of unforeseen events implies an additional incompleteness of the contract. If an unforeseen event does occur, this additional incompleteness will be revealed ex-post.

The point we want to stress here concerns the *form* of the incompleteness that is caused by unforeseen events: it follows immediately from this setting that the presence of unforeseen aspects of the environment increases the degree of *rigidity* of the contract, while the presence of unforeseen aspects of behavior increases the degree of *discretion*.

6. Legal default rules

In most societies, contract laws provide a set of default rules that are intended to complement ("fill the gaps of") private contracts. For example, in the U.S., many such default rules are provided in the Uniform Commercial Code. It has been argued by many law-and-economics scholars that legal default rules allow a society to save on transaction costs. Tightly linked

[29] Since the b'_js are actions of the agent, there are two possibilities, both of which are relevant for us: one is that neither party is aware of these possible actions; the other possibility is that only the principal is not aware of them. Since we assumed that the contract is drafted by the principal, what matters most is the principal's (un)awareness.

[30] This one is motivated by a well-known case in the world of cyclism. At some point in the history of this sport, some cyclists started to get autotransfusions (transfusions of blood to themselves) to enhance their performance. Soon afterwards, the regulations were changed to prohibit this trick.

to this positive theory of legal defaults is the *normative* view that the optimal legal default rules are the ones that the majority (in some sense) of contracting pairs would agree to in the absence of transaction costs. This is the so-called "majoritarian" theory of legal defaults, which has gained prominence in the last two decades (see footnote 1). A thorough investigation of these issues is beyond the scope of this paper. This section has the more limited objective of suggesting, by means of a simple example, that our framework is a natural one to address this type of questions.

Consider a population of $M > 2$ contracting pairs. The contracting problem for each pair is the same as in our basic model of section 3, with one exception: for each k, a fraction $\rho > 1/2$ of the contracting pairs finds the efficient rule to be $C_k : (e_k \rightarrow a_k) \wedge (\neg e_k \rightarrow \neg a_k)$, while the remaining fraction $1-\rho$ prefers the opposite rule, $\bar{C}_k : (e_k \rightarrow \neg a_k) \wedge (\neg e_k \rightarrow a_k)$. In other words, for ρM pairs the incremental gross profit for k is $\pi_k[b(a_k)s(e_k) + (1-b(a_k))(1-s(e_k))]$, and for $(1-\rho)M$ pairs it is $\pi_k[b(a_k)(1-s(e_k)) + (1-b(a_k))s(e_k)]$.[31] We also normalize $A_N = 1$ and assume $p = 1/2$ (μ uniform) and $r = 0$ (no recalling costs) for simplicity.

The timing is as follows: at time zero, a social planner can write a default contract g^D, then each contracting pair $j = 1, ..., M$ writes a (possibly empty) bilateral contract g^j, then the state is realized and all agents take actions. The social planner maximizes the aggregate net surplus minus the cost of writing the default contract. If π^j and C^j denote respectively the gross profit and the cost of the bilateral contract for pair j, and C^D is the cost of writing the default contract, the social planner's problem is

$$\max_{g^D} \sum_{j=1}^{M} [(1-\delta)\pi^j - C^j] - C^D \tag{6.1}$$

Again, we focus on separable contracts (both for the planner and for private pairs), i.e. contracts where each clause deals with a single aspect k.

Next we need to specify in what sense the default contract complements private contracts. Focus on aspect k for contracting pair j. We assume that the default clause is enforced unless the private contract contains a clause regulating aspect k, in which case the private clause replaces

[31] We do not assume anything about the correlation of preferences across k, so in general there will be 2^N groups, each characterized by a distinct first-best mapping, and any of these groups may be empty.

the default clause. Finally, we assume that parties cannot save anything from "recalling" primitive sentences of the default contracts. Recalling economies are possible only internally to a contract.[32]

It is not hard to show that, for each k, there are only three candidate default clauses: (1) the majoritarian contingent clause $C_k : (e_k \to \neg a_k) \wedge (\neg e_k \to a_k)$, (2) the majoritarian rigid clause R_k: $(\top \to a_k)$,[33] and (3) the empty clause. All other default clauses are dominated. The following proposition describes the structure of the optimal default contract, and the subsequent behavior by private contracting pairs. Let $\pi^* \equiv \min\left\{2y(\frac{1}{M} + 2(1-\rho)), \frac{2y}{(2\rho-1)M}\right\}$, where $y = c/(1-\delta)$.

Proposition 4. *Consider problem (6.1). For each k: If $\pi_k > \pi^*$, the social planner writes a C_k default clause; if $\pi^* > \pi_k > \frac{2y}{M}$, the social planner writes a R_k default clause; and if $\pi_k < \frac{2y}{M}$ no default clause is written. The "minority" contracting pairs subsequently replace the default clause if and only if this is C_k and $\pi_k > 2y$.*

The above proposition suggests a qualification to the majoritarian theory of legal defaults as expressed at the beginning of this section. Our results are in line with this theory only for a subset of high-π_k actions. For these actions, the optimal default rules are the ones that the majority of contracting pairs would agree to in the absence of writing costs. And once the legal default is in place, the "minority" pairs negotiate around it if the issue is important enough to them (i.e. if π_k is sufficiently high), otherwise they accept the inefficient default. For less important actions, the optimal default contract specifies rigid rules, or no rules at all. Note that the optimal default contract is closer to the "majoritarian" contract the larger is the population (M): this is because the relevant writing cost for the social planner is the per capita writing

[32] A potentially restrictive assumption is that the social planner can offer only a single default contract. A menu of default contracts will in general do better than a single default contract. However, notice that in general it will not be optimal to provide each group of citizens with its preferred contract, as there may be up to 2^N such groups, each with a different preferred contract, and it may be very costly to write 2^N complete contracts.

[33] For $p = \frac{1}{2}$ the two rigid clauses R_k and $R_k^{\neg}$ are payoff-equivalent for every pair. R_k is the majoritarian rigid clause for $p > \frac{1}{2}$.

cost c/M; the lower this cost, the more complete the default contract.[34,35]

Notice that the benefit from the default legal system is nonmonotonic in the writing cost c. If c is equal to zero or prohibitively high, there can be no efficiency gain from legal defaults, therefore the efficiency gain must be maximum for some intermediate value of c.

Our framework can also capture another qualification to the majoritarian theory, that has been expressed informally by Ayres and Gertner (1989): if the minority is less likely to negotiate around the defaults, perhaps because of higher transaction costs, then the optimal default rules may not be the majority ones. Suppose for example that for a given k the majority prefers the simple noncontingent rule R_k, whereas the minority prefers the contingent rule C_k. In this case, it is more costly for the minority to negotiate around a "bad" default rule than it is for the majority: a majority pair must pay c to write their preferred rule, while a minority pair must pay $2c$. In this case, one can easily find parameter values for which the optimal default rule is the minority rule. The same can happen if the writing cost c is heterogeous across the population. Suppose for example that the population is partitioned in two groups, and the minority has a higher writing cost c. Then it may well be the case that the optimal default rule is the minority one. Finally, the majoritarian principle may not apply if the π_k weights are heterogeous across the population. Suppose that the minority feels more strongly than the majority about aspect k (i.e. has a higher π_k). If c is such that it is optimal to write a default rule but it is too costly for private pairs to replace it, again it is possible that the optimal default rule is the minority rule.

[34]Note that the optimal default contract *as a whole* may not be the contract that the majority would agree to in the absence of writing costs. In fact, it is possible that the default contract does not coincide with *any* pair's first best contract. This is due to the multi-dimensionality of the problem. The "majority" principle applies for each separate dimension, but it may not apply to the contract as a whole.

[35]We assumed $p = 1/2$, i.e. maximal uncertainty. If uncertainty is lower ($p > 1/2$), results tend to change in the direction of more rigidity in both the legal defaults and the bilateral contracts. If M is large enough, or y is small enough, the "majoritarian" result still obtains, with the only difference that, for a given k, minority pairs may replace the default rule with a rigid rule, if π_k is in some intermediate range.

7. Concluding remarks

We developed a multi-task, principal-agent model of contract incompleteness where rigidity and discretion arise endogenously from the costs of describing the external environment and the agent's behavior. In this concluding section we briefly discuss another potential application of our way of modeling complexity costs.

Although we chose to focus mainly on a setting characterized by conflict of interests between principal and agent, our approach is potentially useful also for a different type of setting, where the key problem is not one of incentives, but rather one of efficient communication of information. This could be the case in situations where a scientific authority issues directives for practitioners (e.g., the U.S. Center for Disease Control issuing protocols for doctors and nurses on how to diagnose or treat a certain disease), or when the head of a large organization issues protocols for lower-level employees (e.g., the U.S. Postal Service issuing instructions for local postal offices on how to process and handle mail under various contingencies), or in employment relationships where the main reason to instruct the agent is that the principal has better information (this could apply to the baby-sitting case). In situations of this kind, the presence of complexity costs may lead to rigidity and/or discretion in the set of instructions communicated by the principal.

To exemplify how this type of setting can be captured with our framework, consider a simple variant of our basic model of section 2: suppose that the interests of the principal and the agent are aligned, but the principal is better informed than the agent on the relevant parameters of the payoff functions. Then, if the principal leaves discretion to the agent, there will be a positive probability (from the point of view of the principal) that the agent will take "wrong" actions. If this probability is relatively high, then discretion implies a larger expected loss of surplus than rigidity, hence the qualitative results are likely to be the same as in our basic model. We are able to prove this rigorously in the extreme case where the agent chooses at random within the set of behaviors that do not violate the principal's instructions. Intuitively, discretion (for a given task) in this case implies that the agent will take the wrong action with 50% probability, therefore leaving discretion implies a larger expected loss of surplus than giving a rigid instruction (see the appendix). Extending the analysis to a more general setting with asymmetric information is an ambitious task, and will have to await future research.

8. Appendix

Proof of Remark 1:

To prove point (i), define for any fixed contract g

$$V(g) = \sum_{s\in S} \mu(s) \left[\max_{b\in BR^g(s)} \pi(s,b) - \min_{b\in B^g(s)} \delta(s,b) \right] - \overline{U}.$$

If there are no writing costs, V^g is the value of any behaviorally equivalent contract satisfying the participation constraint as an equality. If there are writing costs, $V(g)$ is an upper bound on the expected profit of the principal if he wants to prescribe behavior according to correspondence $B^g(\cdot)$. Since there are finitely many correspondences from S to $\{0,1\}^{\Pi^a}$ the set $\mathcal{V} = \{V(g) : g \in F\}$ is finite. Any contract $g \in \overline{G}$ (i.e., satisfying the participation constraint as an equality) and such that $V(g) = \max \mathcal{V}$ is a first best contract. For any first best contract there is a behaviorally equivalent, efficiently written contract because the set $\{C : C = C(h)$ for some $h \in F$ behaviorally equivalent to $g\}$ has a minimum (the set is infinite, but it is bounded below by 0 and it is nowhere dense).

The proof that an optimal contract always exists is similar. For every (feasible) behavioral correspondence $B(\cdot)$ we can find an efficiently written contract $g \in \overline{G}$ such that $\forall s \in S$, $B^g(s) = B(s)$. Out of the finite set of contracts obtained in this way we select one that maximizes the net value $V(g) - C(g)$. This contract is optimal.

Points (ii) and (iii) are straightforward. ■

Proof of Remark 2:

Define the finite set $\mathcal{V}$ as above. Let V^1 and V^2 respectively denote the largest and the second to largest element of $\mathcal{V}$. Let

$$C^* = \sup\{C(g) : g \in F,\ g \text{ is efficiently written}\}.$$

Note that $C^* < \infty$ because the set of behavioral correspondences is finite. Since there is at least one efficiently written first best contract whose expected profit is at least $V^1 - C^*$, if $V^1 - C^* > V^2$ then a second best contract must also be a first best.

We now provide an upper bound on the cost efficiently written contracts. Let $\#\Pi^e = E$ and $\#\Pi^a = A$. For any $s \in S$, let $\Pi^e_v(s) = \{q \in \Pi^e : s(q) = v\}$, $v = 0, 1$. $\Pi^a_v(b)$ is similarly defined ($\Pi^e_1(s)$ is the set of primitive sentences about the environment satisfied in state s). Take an efficiently written contract g. Then there is a behaviorally equivalent (possibly inefficient) contract g' with at most 2^E clauses of the form

$$\bigwedge_{s\in S}\left(\left(\left(\bigwedge_{q\in\Pi^e_1(s)} q\right)\wedge\left(\bigwedge_{q\in\Pi^e_0(s)}\neg q\right)\right)\rightarrow\bigvee_{b\in B^g(s)}\left(\left(\bigwedge_{q\in\Pi^a_1(b)} q\right)\wedge\left(\bigwedge_{q\in\Pi^a_0(b)}\neg q\right)\right)\right).$$

Since $r \le c$, cost formula (2.1) implies that

$$C^* \le C(g') \le 2^E\left(E + 2^A \cdot A\right)c.$$

Define

$$\overline{c} = \frac{V^1 - V^2}{2^E(E + 2^A \cdot A)}$$

Then, if $c < \overline{c}$, a second best contract is also a first best contract. ■

Proof of Proposition 1:

Consider an arbitrary contract g^0. We will construct a contract that has the features described in the proposition, yields weakly higher expected gross surplus than g^0 and has a weakly lower writing cost than g^0. We construct this contract in three steps.

1. From g^0 we construct a contract g' which induces a constraint set that is a Cartesian product for each s: $B^{g'}(s) = \prod_{n=1}^N B^{g'}_n(s)$, where $B^{g'}_n(s)$ is the n-th projection of $B^{g'}(s)$. Define the following index sets: $E(g^0) = \{n \in N : e_n \text{ occurs in } g^0\}$, $A(g^0) = \{n \in N : a_n \text{ occurs in } g^0\}$, $E_1(s, g^0) = \{k \in E(g^0) : s(e_k) = 1\}$, $E_0(s, g^0) = \{\ell \in E(g^0) : s(e_\ell) = 0\}$, $A_1(b, g^0) = \{m \in A(g^0) : b(a_m) = 1\}$, $A_0(b, g^0) = \{n \in A(g^0) : b(a_n) = 0\}$. The following is a logically equivalent formulation of g^0 :

$$\bigwedge_{s\in S}\left(\left(\bigwedge_{k\in E_1(s,g^0)} e_k\right)\wedge\left(\bigwedge_{\ell\in E_0(s,g^0)}\neg e_\ell\right)\rightarrow\bigvee_{b\in B^{g^0}(s)}\left(\left(\bigwedge_{m\in A_1(b,g^0)} a_m\right)\wedge\left(\bigwedge_{n\in A_0(b,g^0)}\neg a_n\right)\right)\right)$$

(as usual, conjunctions ranging over empty sets should be replaced by $\top$). Now consider the following contract:

$$g' = \bigwedge_{s\in S}\left(\left(\bigwedge_{k\in E_1(s,g^0)} e_k\right)\wedge\left(\bigwedge_{\ell\in E_0(s,g^0)}\neg e_\ell\right)\rightarrow\left(\bigwedge_{m\in A_1^*(s,g^0)} a_m\right)\wedge\left(\bigwedge_{n\in A_0^*(s,g^0)}\neg a_n\right)\right)$$

where $A_0^*(s,g^0) = \{n : e_n$ and a_n occur in g^0 more than once, and $s(e_n) = 0\}$, and $A_1^*(s,g^0) = A(g^0)\backslash A_0^*(s,g^0)$.

We argue that contract g' yields a higher expected surplus than g^0. Since the agent minimizes the gross surplus, under both contracts and for each a_n not contemplated in g^0, he chooses to "mismatch", i.e. he chooses $b(a_n) = 1 - s(e_n)$, which yields zero incremental gross surplus. Therefore we only have to compare the agent's behavior under g^0 and g' for actions a_n contemplated in g^0. If e_n and a_n occur in g^0 more than once, g' forces the agent to take the right action (a_n or $\neg a_n$) in all states, so the incremental gross surplus for aspect n is maximum. If e_n or a_n occur in g^0 at most once, g' forces the agent to take action a_n in all states. This yields expected incremental gross surplus $pA_N\pi_n$, which is an upper bound to what can be achieved by including e_n (or $\neg e_n$) at most once in the contract.

Therefore we have $V(g') \geq V(g^0)$.

2. From g' we will construct a contract g^* that is separable in the N dimensions. But first we introduce some convenient notation. For each k, s and b, let $s_k = s(e_k)$, $b_k = b(a_k)$. Let $s_{-k} = (s_1, ..., s_{k-1}, s_{k+1}, ..., s_N)$ and (s'_k, s_{-k})=$(s_1, ..., s_{k-1}, s'_k, s_{k+1}, ..., s_N)$. The marginal probability of s_k is denoted $\mu_k(s_k)$. Note that by assumptions (3.1), (3.2) and (3.3), the best response of the agent to state s given contract g is uniquely determined:

$$BR^g(s) = \arg\min_{b\in B^g(s)} \pi(s,b).$$

The k^{th} coordinate of this function is denoted by $BR_k^g(s)$, that is, $BR_k^g(s) = 1$ if under contract g the agent chooses a_k at state s and $BR_k^g(s) = 0$ if the agent chooses $\neg a_k$ at state s. Also, let

$$\sigma_k(s_k, b_k) = (1-\delta)A_N\pi_k[s_k b_k + (1-s_k)(1-b_k)]$$

denote the k^{th} term of the gross surplus.

Finally, recall the notation: $C_k = (e_k \to a_k) \wedge (\neg e_k \to \neg a_k)$, $R_k = \top \to a_k$ and $D = \top \to \top$.

For each $k = 1, ..., N$, pick

$$s_{-k}^* \in \arg\max_{s_{-k}} \left[\sum_{s_k \in \{0,1\}} \mu_k(s_k) \sigma_k(s_k, BR_k^{g'}(s_k, s_{-k})) \right].$$

We construct g^* in the following way: $g^* = \bigwedge_{k=1}^N \gamma_k^*$, where

$$\gamma_k^* = \begin{cases} C_k & \text{if either } \left(B_k^{g'}(0, s_{-k}^*) = \{0\}, B_k^{g'}(1, s_{-k}^*) = \{1\} \right) \\ & \text{or } \left(B_k^{g'}(0, s_{-k}^*) = \{1\}, B_k^{g'}(1, s_{-k}^*) = \{0\} \right) \\ D & \text{if } B_k^{g'}(0, s_{-k}^*) = B_k^{g'}(1, s_{-k}^*) = \{0, 1\} \\ R_k & \text{otherwise} \end{cases}$$

(Recall that $B_k^{g'}(s)$ denotes the k^{th} projection of the constraint set $B^{g'}(s)$.) We argue that g^* yields a weakly higher expected gross surplus than g'. Note that, by definition of g^*, $BR_k^{g*}(s_k, s_{-k})$ is independent of s_{-k}; thus it makes sense to write $BR_k^{g*}(s_k)$. By definition of g^*, additive separability, independence and conflict of interests, we have

$$V(g') = \sum_s \prod_{n=1}^N \mu_n(s_n) \left[\sum_{k=1}^N \sigma_k(s_k, BR_k^{g'}(s)) \right] =$$

$$\sum_{k=1}^N \left\{ \sum_{s_{-k}} \prod_{n \neq k} \mu_n(s_n) \sum_{s_k \in \{0,1\}} \mu_k(s_k) \sigma_k(s_k, BR_k^g(s_k, s_{-k})) \right\} \leq$$

$$\sum_{k=1}^N \arg\max_{s_{-k}} \left[\sum_{s_k \in \{0,1\}} \mu_k(s_k) \sigma_k(s_k, BR_k^{g'}(s_k, s_{-k})) \right] \leq$$

$$\sum_{k=1}^N \sum_{s_k \in \{0,1\}} \mu_k(s_k) \sigma_k(s_k, BR_k^{g*}(s_k)) = V(g^*).$$

Therefore $V(g^*) \geq V(g') \geq V(g^0)$.

Next we argue that $C(g^*) \leq C(g^0)$. We show that, for each aspect $k \in N$, the incremental cost of clause γ_k^* is weakly lower than the incremental cost due to the occurrences of the primitive sentences e_k and a_k in contract g^0. For each k such that $\gamma_k^* = D$, the incremental cost for g^* is zero, which is the minimum. For each k such that $\gamma_k^* = R_k$, primitive sentence a_k must occur at least once in g'. By definition of g', this implies that a_k occurs at least once in g^0. Therefore the incremental cost for g^* is weakly lower than the corresponding one for g^0. For each k such that $\gamma_k^* = C_k$, both a_n and e_n must occur at least twice in g'. By definition of g', this implies that they occur at least twice in g^0, so again the incremental cost for g^* is weakly lower than the corresponding one for g^0. We can conclude that contract g^* is weakly less costly than g^0. Therefore $V(g^*) - C(g^*) \geq V(g^0) - C(g^0)$.

3. We have thus far shown that there is no loss of generality in restricting attention to separable contracts where each clause k is one of the three candidates: C_k, R_k or D. The last step is to determine which of these is optimal to include in the contract for each k. This depends on the parameters p, δ, c and π_k. Since $p > \frac{1}{2}$, the threshold values for π_k are ordered as follows

$$\frac{1}{pA_N}\frac{c}{1-\delta} < \frac{1}{(1-p)A_N}\frac{c+2r}{1-\delta},$$

where D is optimal for $\pi_k < \frac{1}{pA_N}\frac{c}{1-\delta}$, R_k is optimal for $\frac{1}{pA_N}\frac{c}{1-\delta} < \pi_k < \frac{1}{(1-p)A_N}\frac{c+2r}{1-\delta}$, and C_k is optimal for $\pi_k > \frac{1}{(1-p)A_N}\frac{c+2r}{1-\delta}$. Taking into account that π_k is decreasing in k we obtain that the contract stated in the proposition is optimal. Our genericity assumption (3.5) implies that any optimal contract must be equivalent to this one. ■

Proof of Corollary 1: Parts (i) and (ii) are immediate consequences of Proposition 1.

(iii) Suppose that $y < p\frac{\pi_2}{\pi_1+\pi_2}$. Let the optimal contract be denoted by $g(N)$. Then Proposition 1 implies that $g(1)$ contains one clause and $g(2)$ contains two clauses. An increase from N to $N+1$ has two effects: on the one hand, there is one more "gross-benefit opportunity", on the other hand each "gross benefit opportunity" yields a lower gross benefit ($A_{N+1} \leq A_N$), because the total potential gross benefit is independent of N. Let $K(N)$ denote the number of clauses in contract $g(N)$, i.e., $K(N) = \max\{k \in N : pA_N\pi_k > y\}$. Note that for all k and N

$$pA_{N+1}\pi_{k+1} < pA_{N+1}\pi_k < pA_N\pi_k$$

There are three possibilities (assumption (3.5) rules out knife-edge cases):

(1) if $y < pA_{N+1}\pi_{K(N)+1}$, then we have a "corner solution" and $K(N+1) = N+1 = K(N)+1$;

(2) if $pA_{N+1}\pi_{K(N)+1} < y < pA_{N+1}\pi_{K(N)}$, then $K(N) = K(N+1)$;

(3) if $pA_{N+1}\pi_{K(N)} < y$, then $K(N+1) < K(N)$.

As N increases we go from case (1) to case (2) to case (3). ■

Proof of Remark 4

Using techniques similar to parts 1 and 2 of the proof of Proposition 1, one can show that it is possible to maximize task by task. Then the claim is implied by the following observations:

1. The cost of a contingent clause is at least $2c$, the cost of a rigid clause is c and the cost of a discretionary clause is zero.

2. The benefit of a contingent clause is at most equal to

$$\pi_n \left(\sum_{s\in E_n^*} g_n(1,s)\mu(s) + \sum_{s\in \bar{E}_n^*} g_n(0,s)\mu(s) \right) \equiv \pi_n G_n^{FB},$$

where E_n^* is the set of states where it is efficient to execute a_n and $\bar{E}_n^* = S \backslash E_n^*$ is its complement.

3. The benefit of a discretionary clause is

$$\pi_n \left(\sum_{s\in E_n^*} g_n(0,s)\mu(s) + \sum_{s\in \bar{E}_n^*} g_n(1,s)\mu(s) \right) \equiv \pi_n G_n^{D}.$$

4. The benefit of a rigid clause is

$$\pi_n \max \left\{ \sum_{s\in E_n^*} g_n(1,s)\mu(s) + \sum_{s\in \bar{E}_n^*} g_n(1,s)\mu(s), \sum_{s\in E_n^*} g_n(0,s)\mu(s) + \sum_{s\in \bar{E}_n^*} g_n(0,s)\mu(s) \right\} \equiv \pi_n G_n^{R}.$$

5. $G_n^D < G_n^R < G_n^{FB}$ and $G_n^R > \frac{1}{2}(G_n^{FB} + G_n^D)$;

6. The critical value of π_n for which a rigid clause is equivalent to an empty clause is $\pi_n^{R/D} = \frac{c}{G_n^R - G_n^D}$;

7. For a contingent clause to be preferred to a rigid clause it must be $\pi_n > \pi_n^{R/C} = \frac{c}{G_n^{FB} - G_n^R}$;

8. Finally note that $\pi_n^{R/C} > \pi_n^{R/D}$, which implies the claim. ■

Proof of Remark 5

Let us look at a single task a_n. Using observations 1-5 in the previous proof, it is easy to conclude that, as y increases, the optimal clause for task n switches from contingent, to rigid, to discretionary. Aggregating over the N tasks, the claim follows immediately. ■

Proof of Proposition 2

Solving for the optimal contract boils down to finding the optimal number of tasks to be included in the contract. Let $M \in \{0, 1, 2\}$ denote this number. In the additively separable benchmark case, the optimum is $M^{sep} = \max\{k : \pi_k - \delta \geq c\}$, that is, we include in the contract those tasks whose "marginal benefit" at least equals the marginal writing cost. Clearly, the optimal M is given by

$$M^{sep} = \begin{cases} 2 & if \quad c \leq \pi_2 - \delta \\ 1 & if \quad \pi_2 - \delta < c \leq \pi_1 - \delta \\ 0 & if \quad c > \pi_1 - \delta \end{cases}$$

In the case of supermodular $\pi(b)$, the optimum depends on the marginal benefit function MB_k, where

$$MB_1 = g(\pi_1, 0) - \delta, \qquad MB_2 = g(\pi_1, \pi_2) - g(\pi_1, 0) - \delta$$

MB_k may be increasing or decreasing in k. If MB_k is increasing, the solution is bang-bang, in the sense that the optimal contract is either complete or empty. In particular, the optimal number of tasks specified in the contract is depending on whether c is below or above the critical level $\frac{1}{2}(\pi_1 + \pi_2) - \delta$:

$$M^{com} = \begin{cases} 2 & if \quad c < \frac{1}{2}(\pi_1 + \pi_2) - \delta \\ 0 & if \quad c > \frac{1}{2}(\pi_1 + \pi_2) - \delta \end{cases}$$

It follows that task complementarities make the contract (weakly) less incomplete if $c < \frac{1}{2}(\pi_1 + \pi_2) - \delta$ and (weakly) more incomplete if $c > \frac{1}{2}(\pi_1 + \pi_2) - \delta$.

If MB_k is decreasing, then $M^{com} = \max\{k : MB_k \geq c\}$. It is easy to show that, given the assumptions made on the g function, we have $MB_1 < \pi_1 - \delta$ and $MB_2 > \pi_2 - \delta$. This implies that for $\pi_2 - \delta < c < MB_2$ the contract is less incomplete under supermodular $\pi(b)$ than under separable $\pi(b)$, for $MB_1 < c < \pi_1 - \delta$ the opposite is true, and for all other values of c the contract is equally incomplete in the two cases. ■

Proof of Proposition 3

Again, it can be shown that the optimal contract can be derived by focusing separately on each elementary action. To simplify notation, we will pretend that there is only one elementary action, a. The corresponding payoff can then be written as $\pi = g(s, b)$, where $s = (s_1, ..., s_m) \in \{0, 1\}^m$, $b \in \{0, 1\}$, $s_k = 1$ means that e_k occurs, and $b = 1$ means that a is chosen.

We identify an incentive contract $(\varphi_k \to t_k)_{k=1}^K$with the corresponding transfer function $T : S \times \{0, 1\} \to \mathbf{R}$, where $T(s, b) = \sum_{k:(s,b)\in\|\varphi_k\|} t_k$. For any incentive contract $T(s, b)$, we can focus on the difference $\Delta_T(s) = T(s, 1) - T(s, 0)$ which is the monetary incentive to take action a at state s. The cost of incentive contract T is at least $c\,(1 + n(T))$ where $n(T)$ is the number of elementary events affecting the value of $\Delta_T(s)$:

$$n(T) = \#\left\{k : \exists s_{-k}, \Delta_T(1, s_{-k}) \neq \Delta_T(0, s_{-k})\right\},$$

(recall that $\#X$ denotes the cardinality of set X).

An incentive contract T *implements* a behavioral function $\mathbf{b}_T : S \to \{0, 1\}$ if, for all s and b,[36]

$$\begin{aligned}\mathbf{b}_T(s) &= 1 \Leftrightarrow \Delta_T(s) - \delta\pi(s, 1) > -\delta\pi(s, 0),\\ \mathbf{b}_T(s) &= 0 \Leftrightarrow \Delta_T(s) - \delta\pi(s, 1) < -\delta\pi(s, 0).\end{aligned}$$

Let

$$E^+(T) = \{s : \Delta_T(s) > 0\}$$

be the set of states where T gives a monetary incentive to choose a. Futhermore, let

$$K(T) = \{k : \exists s_{-k}, \Delta_T(1, s_{-k}) > 0 \geq \Delta_T(0, s_{-k}), \text{ or } \Delta_T(1, s_{-k}) \leq 0 < \Delta_T(0, s_{-k})\}$$

be the set of indexes of elementary variables which are relevant to determine whether $s \in E^+(T)$ or not. We let $s_{K(T)}$ denote a typical element of the "relevant state space" $\{0, 1\}^{K(T)}$ and we

[36]Note that this is a notion of strict implementation. We restrict our attention to strict implementation only for the sake of simplicity.

write $s = (s_{K(T)}, s_{-K(T)})$, where $-K(T)$ is the complementary set of indexes. Clearly $E^{+}(T)$ is just the product of its projections on $\{0,1\}^{K(T)}$ and $\{0,1\}^{-K(T)}$:

$$E^{+}(T) = E^{+}_{K(T)} \times \{0,1\}^{-K(T)},$$

where

$$E^{+}_{K(T)} = \left\{ s_{K(T)} : \exists s_{K(T)}, (s_{K(T)}, s_{-K(T)}) \in E^{+}(T) \right\}.$$

Note that the cardinality of $K(T)$ is not larger than $n(T)$.

Now we can write a conjunctive-disjunctive formula $\eta^{+}(T)$ whose truth set is $E^{+}(T)$:

$$\eta^{+}(T) = \bigvee_{s_{K(T)} \in E^{+}(T)} \left(\bigwedge_{k \in K(T): s_k = 1} e_k \right) \wedge \left(\bigwedge_{k \in K(T): s_k = 0} \neg e_k \right).$$

Clearly, the cost of $\eta^{+}(T)$ is $c \cdot \# K(T)$ (recall that we assumed $r = 0$).

Let $\mathbf{b}_f$ denote the behavioral function implemented by a forcing contract f, where

$$\mathbf{b}_f(s) = \arg \min_{b \in B^f(s)} \pi(s, b)$$

We are now ready to prove the result. Formally, our claim is the following. Fix an arbitrary incentive contract T and suppose that T implements the behavioral function $\mathbf{b}_T : S \to \{0,1\}$. Then the forcing contract $f = (\eta^{+}(T) \to a) \wedge (\neg \eta^{+}(T) \to \neg a)$ weakly dominates T, that is,

$$C(f) = c(1 + \# K(T)) \leq c(1 + n(T)) \leq C(T)$$

and

$$\forall s \in S,\ g(s, \mathbf{b}_f(s)) \geq g(s, \mathbf{b}_T(s)).$$

To prove the claim, first note that, since $\# K(T) \leq n(T)$, we have $C(f) \leq C(T)$.

Since T implements $\mathbf{b}_T$, we have

$$\begin{aligned} \mathbf{b}_T(s) &= 1 \Leftrightarrow g(s,0) - g(s,1) > -\frac{\Delta_T(s)}{\delta}, \\ \mathbf{b}_T(s) &= 0 \Leftrightarrow g(s,1) - g(s,0) > \frac{\Delta_T(s)}{\delta}. \end{aligned}$$

The behavioral function $\mathbf{b}_f$ implemented by the forcing contract f satisfies

$$\begin{aligned}\mathbf{b}_f(s) &= 1 \Leftrightarrow \frac{\Delta_T(s)}{\delta} > 0, \\ \mathbf{b}_f(s) &= 0 \Leftrightarrow -\frac{\Delta_T(s)}{\delta} \geq 0.\end{aligned}$$

Therefore

$$g(s, \mathbf{b}_f(s)) - g(s, \mathbf{b}_T(s)) = \begin{cases} g(s,1) - g(s,0), & \text{if } g(s,1) - g(s,0) > \frac{\Delta_T(s)}{\delta} > 0, \\ g(s,0) - g(s,1), & \text{if } g(s,0) - g(s,1) > -\frac{\Delta_T(s)}{\delta} \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

The claim follows. ■

Proof of Remark 7:

Suppose a language has Q primitive sentences, $p_1, p_2, ..., p_Q$. Define an *atom* of this language to be a formula of the form $q_1 \wedge q_2 \wedge ... \wedge q_Q$, where each q_j is either p_j or the negation of p_j.

A language with Q elementary sentences ha 2^Q atoms. These atoms have disjoint truth sets, some of which may be empty. Let A_0 be the number of atoms with nonempty truth sets. Of course we have $A_0 \leq 2^Q$. Now note that the set of of events that this language can describe consists of all the unions of atoms in A_0. The cardinality of this set – i.e., the expressive power – is 2^{A_0} (minus one, if the impossible event is subtracted). A natural language with Q elementary sentences has all nonempty atoms, therefore its expressive power is 2^{2^Q}.

Let $\mathcal{L}_R$ be a language with R elementary sentences and let $\mathcal{L}_Q^*$ be a natural language with Q elementary sentences. If $R \leq Q$ then the expressive power of $\mathcal{L}_R$ is at most $2^{2^R} \leq 2^{2^Q}$. If $\mathcal{L}_R$ has the same expressive power as $\mathcal{L}_Q^*$, then $2^{2^R} \geq 2^{2^Q}$, i.e., $R \geq Q$. ■

Proof of Proposition 4:

Let us focus on aspect k. It is not hard to show that only the empty clause D, the rigid clause R_k and the (double) contingent clause C_k can be optimal default clauses (the rigid rules R_k and R_k^- are payoff-equivalent when $p = \frac{1}{2}$; without loss of generality we consider only R_k). Note that, since $p = \frac{1}{2}$, it is never optimal for private parties to include a rigid rule in the contract, but it can be optimal not to contract around a rigid default rule. The following table shows when it is optimal for the "majority" to replace the default clause with C_k. We put in parentheses in the third column the private incremental payoff when the condition in the second column is not satisfied and hence the default clause is not replaced. The analysis for the "minority" is analogous.

Default Clause k	Replaced with C_k if	Private Incremental Payoff for k
D	$\pi_k \geq 2y$	$(1-\delta)\pi_k - 2c$ (or 0)
C_k	never	$(1-\delta)\pi_k$
R_k	$\pi_k \geq 4y$	$(1-\delta)\pi_k - 2c$ (or $\frac{1}{2}(1-\delta)\pi_k$)

The next table summarizes the planner's incremental payoff from each default clause as a function of the parameters.

Default	$\pi_k < 2y$	$2y < \pi_k < 4y$	$4y < \pi_k$
D	0	$M\left[(1-\delta)\pi_k - 2c\right]$	$M\left[(1-\delta)\pi_k - 2c\right]$
R_k	$\frac{1}{2}M(1-\delta)\pi_k - c$	$\frac{1}{2}M(1-\delta)\pi_k - c$	$M\left[(1-\delta)\pi_k - 2c\right] - c$
C_k	$\rho M(1-\delta)\pi_k - 2c$	$M\left[(1-\delta)\pi_k - 2(1-\rho)c\right] - 2c$	$M\left[(1-\delta)\pi_k - 2(1-\rho)c\right] - 2c$

Case (i): $\pi_k \leq 2y$.

D is preferred to R_k iff $\frac{2y}{M} > \pi_k$.

R_k is preferred to C_k iff $\frac{2y}{M(2\rho-1)} \geq \pi_k$.

Note that $\frac{2y}{M} < \frac{2y}{M(2\rho-1)}$ because $0 < \rho < 1$. Therefore

- D is optimal iff $0 \leq \pi_k \leq \frac{2y}{M}$,
- R_k is optimal iff $\frac{2y}{M} \leq \pi_k \leq 2y \min\{1, \frac{1}{2\rho-1}\}$,
- C_k is optimal iff $2y \min\{1, \frac{1}{2\rho-1}\} \leq \pi_k \leq 2y$.

Case (ii): $2y \leq \pi_k \leq 4y$.

We only have to compare R_k with C_k.

R_k is preferred to C_k iff $2y\left[\frac{1}{M}+2(1-\rho)\right] \geq \pi_k$.

Note that $2y\left[\frac{1}{M}+2(1-\rho)\right] < 4y$ because $M > 2$ and $\rho > \frac{1}{2}$. Therefore

- R_k is optimal iff $2y \leq \pi_k \leq 2y \max\left\{1, \left[\frac{1}{M}+2(1-\rho)\right]\right\}$,
- C_k is optimal iff $2y \max\left\{1, \left[\frac{1}{M}+2(1-\rho)\right]\right\} \leq \pi_k \leq 4y$.

Case (iii): $\pi_k > 4y$.

By inspection of the table above (and taking into account the restrictions on M and ρ) only the contingent clause C_k is optimal in this case.

Now note that the threshold values are ordered as follows:

- If $M = \frac{1}{2\rho-1}$ then $2y = \frac{2y}{M(2\rho-1)} = 2y\left[\frac{1}{M}+2(1-\rho)\right]$,
- if $M < \frac{1}{2\rho-1}$ then $2y < \frac{2y}{M(2\rho-1)} < 2y\left[\frac{1}{M}+2(1-\rho)\right]$,
- if $M > \frac{1}{2\rho-1}$ then $2y > \frac{2y}{M(2\rho-1)} > 2y\left[\frac{1}{M}+2(1-\rho)\right]$.

Therefore the above results can be summarized as follows:

- D is optimal iff $0 \leq \pi_k \leq \frac{2y}{M}$,
- R_k is optimal iff $\frac{2y}{M} \leq \pi_k \leq 2y \max\left\{\frac{1}{M(2\rho-1)}, \left[\frac{1}{M}+2(1-\rho)\right]\right\}$,
- C_k is optimal iff $2y \max\left\{\frac{1}{M(2\rho-1)}, \left[\frac{1}{M}+2(1-\rho)\right]\right\} \leq \pi_k$.

If R_k is the optimal default, neither the "majority" nor the "minority" contracting pairs replace it. If C_k is the optimal default, the "minority" contracting pairs replace it whenever $\pi_k > 2y$. ■

Characterization of the Optimal Contract with Random Behavior:

Assume that the agent chooses at random in $B^g(s)$. To simplify the analysis also assume that $r = 0$. Let $K_1 = \max\{n : \pi_n > \frac{c}{1-p}\}$, $K_1 + K_2 = \max(K_1, \max\{n : \pi_n > \frac{c}{p-\frac{1}{2}}\})$

Proposition 5. *: Every optimal contract is equivalent to the following:*

$$\bigwedge_{k=1}^{K_1} [(e_k \to a_k) \wedge (\neg e_k \to \neg a_k)] \wedge [\top \to \bigwedge_{k=K_1+1}^{K_1+K_2} a_k]$$

Proof. We initially proceed as in the proof of Proposition 1: we start with an arbitrary contract g^0, then we construct a contract that has the features described in the proposition, yields weakly higher expected gross surplus than g^0 and has a weakly lower writing cost than g^0. We construct this contract in three steps.

1. The following is a logically equivalent formulation of g^0 :

$$\bigwedge_{s \in S} \left(\left(\bigwedge_{k \in E_1(s,g^0)} e_k \right) \wedge \left(\bigwedge_{\ell \in E_0(s,g^0)} \neg e_\ell \right) \to \bigvee_{b \in B^{g^0}(s)} \left(\left(\bigwedge_{m \in A_1(b,g^0)} a_m \right) \wedge \left(\bigwedge_{n \in A_0(b,g^0)} \neg a_n \right) \right) \right)$$

(the proof of Proposition 1). Now consider the following contract:

$$g' = \bigwedge_{s \in S} \left(\left(\bigwedge_{k \in E_1(s,g^0)} e_k \right) \wedge \left(\bigwedge_{\ell \in E_0(s,g^0)} \neg e_\ell \right) \to \left(\bigwedge_{m \in A_1^*(s,g^0)} a_m \right) \wedge \left(\bigwedge_{n \in A_0^*(s,g^0)} \neg a_n \right) \right)$$

where $A_0^*(s, g^0) = \{n : e_n \text{ and } a_n \text{ occur in } g^0 \text{ and } s(e_n) = 0\}$, and $A_1^*(s, g^0) = A(g^0) \backslash A_0^*(s, g^0)$.

We argue that contract g' yields a (weakly) higher expected surplus than g^0. First note that $C(g^0) = C(g')$ because g^0 and g' contain the same set of elementary sentences. Next observe that by additive separability of payoffs we need only compare the expected incremental gross surplus for each aspect n of the contractual problem. Since the agent chooses at random in the state-contingent constraint set, under both contracts, for each a_n not contemplated in g^0 and each state, he chooses a_n with probability 1/2 (which expected incremental gross surplus $\frac{1}{2}\pi_n$). Therefore we only have to compare the agent's behavior under g^0 and g' for elementary actions a_n contemplated in g^0, that is, actions in $A(g^0)$. If e_n and a_n occur in g^0, then g' forces the agent to take the right action (a_n or $\neg a_n$) in all states, so the incremental gross surplus for aspect n is maximum. If a_n occurs in g^0 but e_n does not, then g' forces the agent to take action a_n in all states (note that in this case $a_n \in A_1^*(s, g^0)$). This yields expected incremental gross surplus $p\pi_n$. By the independence assumption, this is an upper bound to what can be achieved without including e_n in the contract. To see this, note that, even accounting for the agent's randomization, if e_n does not appear in g^0, $b(a_n)$ and $s(e_n)$ are independent random variables; therefore the expected incremental surplus from task n in contract g^0 is

$$\begin{aligned} & \pi_n E\left[s(s_n)b(a_n) + (1 - s(e_n)(1 - b(a_n))\right] \\ = \ & \pi_n E[s(e_n)]E[b(a_n)] + \{1 - E(s(e_n)]\}\{1 - E[b(a_n)]\} \\ = \ & \pi_n pE[b(a_n)] + \pi_n(1 - p)\{1 - E[b(a_n)]\} \leq \pi_n p, \end{aligned}$$

where the first equality holds by independence and the inquality follows from $p > \frac{1}{2}$ and $0 \leq E[b(a_n) \leq 1$. (Note that the inequality is strict unless g^0 prescribes a_n in every state.) Therefore we have $V(g') \geq V(g^0)$, where $V(g)$ denotes the net surplus induced by contract g.

2. From g' we will construct a contract g^* that is separable in the N dimensions. As in the proof of Proposition 1, we let $s_{-k} = (s_1, ..., s_{k-1}, s_{k+1}, ..., s_N)$ and (s'_k, s_{-k})=$(s_1, ..., s_{k-1}, s'_k, s_{k+1}, ..., s_N)$. The marginal probability of s_k is denoted $\mu_k(s_k)$. Also, let

$$\sigma_k(s_k, b_k) = \pi_k[s_k b_k + (1 - s_k)(1 - b_k)]$$

denote the k^{th} term of the gross surplus, and let $E^{g'}(\sigma_k|s)$ denote its expected value conditional on s, given contract g'.

For each $k = 1, ..., N$, pick

$$s^*_{-k} \in \arg\max_{s_{-k}} \left\{ \sum_{s_k \in \{0,1\}} \mu_k(s_k) E^{g'}(\sigma_k | s_k, s_{-k}) \right\}.$$

We construct g^* in the following way: $g^* = \bigwedge_{k=1}^N \gamma_k^*$, where

$$\gamma_k^* = \begin{cases} C_k & \text{if} \quad B_k^{g'}(0, s^*_{-k}) \neq B_k^{g'}(1, s^*_{-k}), \\ D & \text{if } B_k^{g'}(0, s^*_{-k}) = B_k^{g'}(1, s^*_{-k}) = \{0, 1\}, \\ R_k & \text{otherwise.} \end{cases}$$

(Recall that $B_k^{g'}(s)$ denotes the k^{th} projection of the constraint set $B^{g'}(s)$.) First note that every elementary sentence contained in g^* is also contained in g' (for example, e_k occurs in g^* only if $\gamma_k^* = C_k$, which implies that $B_k^{g'}(s)$ depends on s_k; thus e_k must also occur in g'). Therefore $C(g^*) \leq C(g')$. Next we argue that g^* yields a weakly higher expected gross surplus than g'. By definition of g^*, additive separability, independence and uniform randomization, we have

$$V(g') + C(g') = \sum_s \prod_{n=1}^{N} \mu_n(s_n) \sum_{k=1}^{N} E^{g'}(\sigma_k|s) =$$

$$\sum_{k=1}^{N} \left\{ \sum_{s_{-k}} \prod_{n \neq k} \mu_n(s_n) \sum_{s_k \in \{0,1\}} \mu_k(s_k) E^{g'}(\sigma_k|s_k, s_{-k}) \right\} \leq$$

$$\sum_{k=1}^{N} \max_{s_{-k}} \left\{ \sum_{s_k \in \{0,1\}} \mu_k(s_k) E^{g'}(\sigma_k|s_k, s_{-k}) \right\} \leq$$

$$\sum_{k \in K_C(g^*)} \pi_k + \sum_{k \in K_R(g^*)} p\pi_k + \sum_{k \in K_D(g^*)} \frac{1}{2}\pi_k = V(g^*) + C(g^*),$$

where $K_C(g^*) = \{k : \gamma_k = C_k\}$, $K_R(g^*) = \{k : \gamma_k = R_k\}$, $K_D(g^*) = \{k : \gamma_k = D\}$. Since $C(g^*) \leq C(g')$, we obtain $V(g^*) \geq V(g')$. Therefore $V(g^*) \geq V(g^0)$.

3. We have thus far shown that there is no loss of generality in restricting attention to separable contracts where each clause k is one of the three candidates: C_k, R_k or D. The last step is to determine which of these is optimal to include in the contract for each k. This depends on the parameters p, c and π_k. If $p > \frac{3}{4}$, the threshold values for π_k are ordered as follows

$$\frac{c}{p - \frac{1}{2}} < \frac{c}{1 - p},$$

where D is optimal for $\pi_k < \frac{c}{p - \frac{1}{2}}$, R_k is optimal for $\frac{c}{p - \frac{1}{2}} < \pi_k < \frac{c}{(1-p)}$, and C_k is optimal for $\pi_k > \frac{c}{(1-p)}$. If $p < \frac{3}{4}$ the rigid clause cannot be optimal. Taking into account that π_k is decreasing in k we obtain the characterization of Proposition 5. ■

References

[1] ALLEN, F. and D. GALE (1992): "Measurement Distortion and Missing Contingencies in Optimal Contracts," *Economic Theory,* **26**, 163-179.

[2] AL NAJJAR, N., L. ANDERLINI and L. FELLI (2001): "Unforeseen Contingencies," mimeo.

[3] ANDERLINI, L. and L. FELLI (1994): "Incomplete Written Contracts: Undescribable States of Nature," *Quarterly Journal of Economics,* **109**, 1085-1124.

[4] ANDERLINI, L. and L. FELLI (1998): "Describability and Agency Problems," *European Economic Review*, **42**, 35-59.

[5] ANDERLINI, L. and L. FELLI (1999): "Incomplete Contracts and Complexity Costs," *Theory and Decision,* **46** (1), 23-50.

[6] ANDERLINI, L. and L. FELLI (1997): "Costly Coasian Contracts," mimeo, University of Cambridge.

[7] AYRES, I. and R. GERTNER (1989): "Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules," *Yale Law Journal*, **99**, 87-130.

[8] AYRES, I. and R. GERTNER (1992): "Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules," *Yale Law Journal*, **99**, 87-130.

[9] BAIRD and JACKSON (1985): "Fraudolent Conveyance Law and Its Proper Domain," *Vanderbilt Law Review*, **38**.

[10] BATTIGALLI, P. and G. MAGGI (2000): "Contracting Over Time when Writing is Costly," Princeton University, Working Papers in Economic Theory (http://www.princeton.edu/~ectheory/00F.html).

[11] BERNHEIM, D. and M. WHINSTON (1998): "Incomplete Contracts and Strategic Ambiguity," *American Economic Review*, **88**, 902-932.

[12] BOOT, A., S. GREENBAUM and A. THAKOR (1993): "Reputation and Discretion in Financial Contracting," *American Economic Review,* **83**, 1165-83.

[13] DEWATRIPONT, M and E. MASKIN (1995): "Contractual Contingencies and Renegotiation," *Rand Journal of Economics*, 26, 704-719.

[14] DOUGLAS, J.A. and J.O'NEILL (1994), *Construction and Development Forms: Planning, Financing, Building*, Boston MA: Warren, Gorham and Lamont.

[15] DYE, R. A. (1985): "Costly Contract Contingencies," *International Economic Review*, **26**, 233-50.

[16] EASTERBROOK and FISCHEL (1985): "Limited Liability and the Corporation," *University of Chicago Law Review,* **52**, 89-102.

[17] FUDENBERG, D. and J. TIROLE (1991): *Game Theory.* Cambridge MA: MIT Press.

[18] GOETZ and SCOTT (1983): "The Mitigation Principle: Toward a General Theory of Contractual Obligation," *Vanderbilt Law Review,* **69**.

[19] HAMILTON, A.G. (1988): *Logic for Mathematicians.* Cambridge UK: Cambridge University Press.

[20] HART, O. (1995): *Firms, Contracts, and Financial Structure.* Oxford: Oxford University Press.

[21] HART, O. and MOORE (1999): "Foundations of Incomplete Contracts," *Review of Economic Studies*, **66**, 115-138.

[22] KRASA, S. and S.R. WILLIAMS (1999): "Incompleteness as a Constraint in Contract Design," mimeo, University of Illinois.

[23] MASKIN, E. and J. TIROLE (1999): "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies*, **66**, 83-114.

[24] MEIHUIZEN, H.E. and S.N.WIGGINS (2000), "Information Cascades and Contractual Incompleteness in Natural Gas Contracting," mimeo.

[25] MILGROM, P and J. ROBERTS (1990), "The Economics of Modern Manufacturing: Technology, Strategy and Organization," *American Economic Review,* **80**, 511-528.

[26] MUKERJI, Sujoy (1998), "Ambiguity aversion and incompleteness of contractual form", *American Economic Review,* **88**, 1207-31.

[27] POSNER, R. (1986): *Economic Analysis of Law,* 3rd ed., Boston: Little Brown.

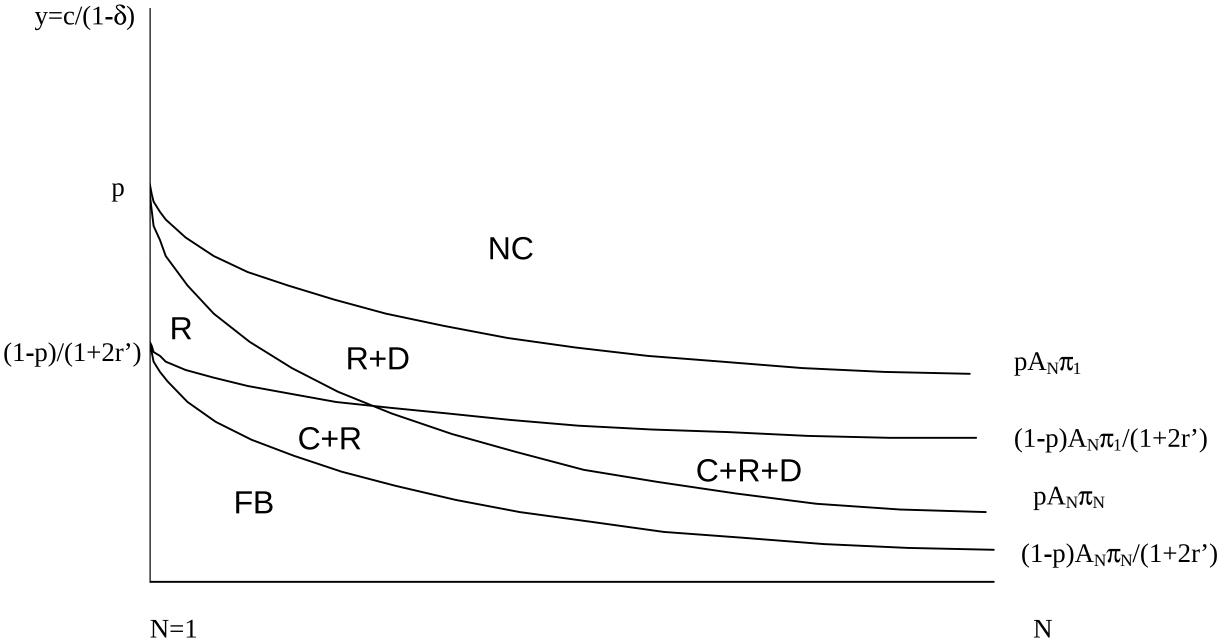
[28] SEGAL, I. (1995): *Essays on Commitment, Renegotiation and Incompleteness of Contracts*, PhD thesis, Harvard University.

[29] SEGAL, I. (1999): "Complexity and Renegotiation: A Foundation for Incomplete Contracts," *Review of Economic Studies*, **66**, 57-82.

[30] SPIER, K. (1992): "Incomplete Contracts and Signalling," *Rand Journal of Economics*, **26**, 163-179.

[31] TIROLE, J. (1999): "Incomplete Contracts: Where Do We Stand?" *Econometrica,* **67,** 741-781.

Figure 1


y=c/(1-δ)
p
(1-p)/(1+2r')
NC
R
R+D
C+R
C+R+D
FB
pA_Nπ_1
(1-p)A_Nπ_1/(1+2r')
pA_Nπ_N
(1-p)A_Nπ_N/(1+2r')
N=1
N