Explicit vs. Implicit Incentives

Margaret A. Meyer
Nuffield College and Department of Economics
Oxford University

2020

Explicit incentives - provided through explicit contractual commitments by principal (P)

Implicit incentives - arise when a principal (e.g. a firm) or principals (e.g. firms competing in a labor market) have some *ex post discretion* how to respond to agent's (A's) performance; A has *implicit incentives* to change performance to influence P's discretionary decision

Two different types of implicit incentives (= two different models of incentives deriving from a concern with "reputation"):

- **Relational contracts (=self-enforcing contracts)**
- **Learning about A's intrinsic characteristics**

Implicit incentives - arise when principal(s) has some *ex post discretion* how to respond to agent's (A's) performance; A has *implicit incentives* to change performance to influence P's discretionary decision

Applications of models of implicit incentives:

- incentives for employees (within firms)
- incentives for suppliers (between firms or between firms and consumers)
- incentives for public servants or for expert advisors
- incentives for countries to cooperate and avoid conflict
- more generally, incentives for actors to behave in a manner different from that which is in their short-run self-interest

N.B.: Implicit incentives may *encourage* or *discourage* "cooperative" (efficient) behavior

Relational contracts: this lecture

Learning about intrinsic characteristics: next lecture

Both lectures will

- first explore pure implicit incentives
- then study how implicit and explicit incentives interact

Relational contracts (=self-enforcing contracts)

- informal agreements sustained by the "shadow of the future"
- infinitely-repeated interaction (at least potentially)
- select one equilibrium among many
- can work even if there is no learning about A's intrinsic characteristics

Earliest and most influential applications to:

- incentives for employees: MacLeod and Malcomson, *Etrica,* 1989, and *AER,* 1998; Levin, *AER,* 2003
- incentives for suppliers: Klein and Leffler, *JPE*, 1981; Baker et al, *QJE*, 2002

Overviews: Malcomson, *Handbook of Orgl. Econ.*, 2013; MacLeod, *JEL*, 2007

Outline of presentation of relational contracts:

1. Baker, Gibbons, and Murphy, *QJE*, 1994, "Subjective Performance Measures in Optimal Incentive Contracts" (henceforth, "BGM (1994)")

a. Simple model of "pure" implicit incentives for effort

b. Simple model of "pure" but distortionary explicit incentives

c. Interaction between implicit incentives (relational contracts) and explicit incentives (formal contracts)

2. Levin, *AER*, 2003, "Relational Incentive Contracts": provides a more general analysis of relational contracts under hidden action

- more general forms of contracts, production technologies, and informational frictions

3. Dynamics within relational contracts

4. Building effective relational contracts

Simple model of "pure" implicit incentives for effort (BGM, 1994)

- risk-neutral agent A privately chooses e each period at cost $c(e)$
- e influences A's contribution to firm value, $y \in \{0, 1\}$: $P(y = 1 \mid e) = e$
- y not verifiable (objectively measurable), so can't be used in an *explicit* contract
- y can be observed by both firm and A and used in an *implicit* contract
- A has reservation utility $w_a \geq 0$: w_a may reflect intensity of competition for A
- firm has discount rate r (A's discount rate is irrelevant)
- if payment equals I, firm's payoff $= y - I$, and A's payoff $= I - c(e)$
- first-best benchmark: if effort contractible, optimal e^* satisfies $1 = c'(e^*)$
- **stationary contract**: (s, b), where s is (unrestricted) base salary paid if A turns up and $b \geq 0$ is bonus to be paid when $y = 1$; s is explicit component (since whether A turns up is verifiable) and b is implicit component

Simple model of "pure" implicit incentives for effort (BGM, 1994)

Questions: i) Under what conditions will the firm be willing to pay bonus when $y = 1$, i.e. under what conditions will the implicit contract be "self-enforcing"? **ii)** What is the firm's profit-maximizing self-enforcing bonus?

- If A trusts firm to pay bonus when $y = 1$, then A chooses e s.t. $b = c'(e)$.
- Given b, and e induced if A trusts, firm will choose lowest s that induces A to participate ($= s$ s.t. A's IR constraint binds).
- If there were a known finite horizon, relationship would "unravel": firm would never pay bonus, so A would never supply effort.
- Consider an infinitely-repeated relationship and *cooperation* supported by *trigger strategies*:
  - Firm and worker begin by cooperating (i.e. worker chooses e s.t. $c'(e) = b$, and firm pays b iff $y = 1$). Cooperation continues unless firm ever defects. Following a defection, worker chooses $e = 0$ and firm never pays b: these continuation strategies are best responses to each other.

Simple model of "pure" implicit incentives for effort (BGM, 1994)

i) Under what conditions will firm be willing to pay bonus?

Define $V(b)$ = firm's expected per-pd. profit in cooperative eqm., given that s is set so A's IR constraint binds: $V(b) = \hat{e}(b) - c(\hat{e}(b)) - w_a$, where $\hat{e}(b)$ solves A's FOC $b = c'(e)$.

Then the "no reneging constraint" (NRC) for firm is:

$$1 - s - b + \frac{V(b)}{r} \geq 1 - s \iff V(b) \geq rb.$$

ii) What is the profit-maximizing self-enforcing bonus?

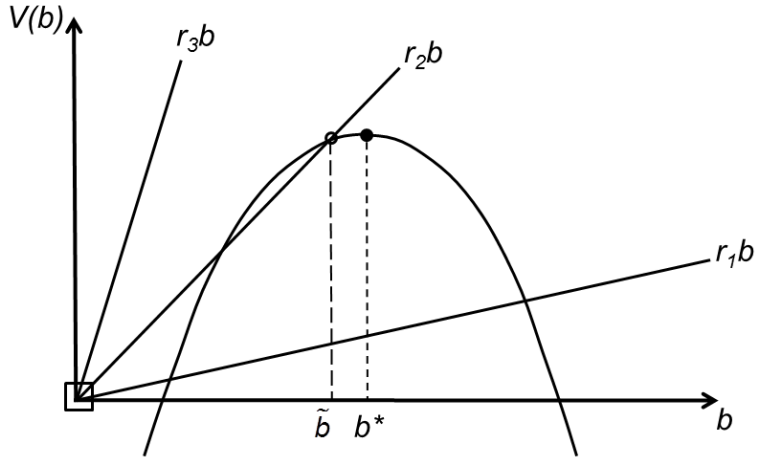
Firm optimally chooses b to solve max $V(b)$ s.t $V(b) \geq rb$.

Explicit-contract benchmark: Let b^* denote the optimal bonus in an explicit contract, i.e. if y were contractible. With y contractible and A risk-neutral, optimal contract implements the first-best effort, which solves $1 = c'(e^*)$. Hence optimal explicit bonus is $b^* = 1$.

Simple model of "pure" implicit incentives for effort (BGM, 1994)

Proposition: *The profit-maximizing self-enforcing bonus* $\widetilde{b}$ *is weakly decreasing in* r *and* w_a*. If* $b = 1$ *satisfies the NRC, then* $\widetilde{b} = b^* = 1$*. If* $b = 1$ *violates the NRC, then* $\widetilde{b}$ *is the largest* b *satisfying the NRC if such a* $b \geq 0$ *exists, otherwise is 0.*

r small (e.g. r_1): $\widetilde{b} = b^*$ is feasible
r medium (e.g. r_2): $0 < \widetilde{b} < b^*$
r large (e.g. r_3): $\widetilde{b} = 0$, since no $b \geq 0$ satisfies the NRC



NB: Diagram assumes $w_a > 0$. Hence $V(0) < 0$.

Simple model of "pure" implicit incentives for effort: Discussion

- In this model, only the firm is concerned with its reputation, i.e. with the present value of the ongoing relationship $V(b)/r$. A's incentives are provided entirely by the short-term (implicit) contract.

- The "firm" may have an incentive to honor the implicit contract even if the individuals making decisions about paying the bonus have *finite* employment horizons—if these indivs. own shares in the firm and can sell their shares to others (e.g. law partnerships), then reneging today reduces value of shares.

  – This observation is due to Kreps (1990) and Cremer (1986).

- Relational contracts can also work even if the firm employs a sequence of agents, each with a finite employment duration— as long as each new agent learns whether the firm ever reneged in the past.

Simple model of "pure" but distortionary explicit incentives (BGM (94))

- A's privately-observed effort stochastically determines a performance measure p, which is objectively measurable, so can be used in an explicit contract.

- But $p \in \{0, 1\}$ is an **imperfect** proxy for y—p and y are independent, conditional on e—**and** A observes the ease of influencing p **before** choosing effort:
  - $P(p = 1 \mid e) = ue$, where u is ex ante random, $E(u) = 1$, and u is observed by A *before* choosing e at cost $c(e) = \gamma e^2$.

- Explicit contract: (s, β), where bonus β is paid iff $p = 1$.

- First-best effort satisfies $1 = c'(e^*)$, so $e^* = (1/2\gamma)$.

- Given (s, β) and u, A chooses e s.t. $u\beta = c'(e)$, so $e = u\beta/2\gamma$.
  - The explicit contract here induces A to engage in **gaming**, i.e. to choose effort that fluctuates with u, even though first-best effort e^* is indep. of u.
  - Fluctuations in effort *per se* reduce A's expected utility (since $c(e)$ is strictly convex), thus increasing the salary s required to induce A to participate.

Simple model of "pure" but distortionary explicit incentives (BGM (94))

Optimal explicit contract: Firm chooses (β, s) to solve

$$\max E_u[e - \beta u e - s]$$

$$\text{subj. to } E_u[s + \beta u e - \gamma e^2] \geq w_a \ (IR) \quad \text{and} \quad e = u\beta/2\gamma \ (IC)$$

Given β, optimal s makes (IR) bind.

Proposition: *The optimal explicit bonus, β^*, satisfies $\beta^* = 1/(1 + var(u)) < 1$ and is decreasing in $var(u)$ and independent of w_a. The firm's maximized per-period profit, $V(\beta^*)$, is decreasing in $var(u)$ and in w_a and can be positive or negative (given that $w_a \geq 0$).*

- Firm trades off strength of incentives against magnitude of (inefficient) fluctuations in effort. The more *distortionary* is the explicit performance measure (the larger is $var(u)$), the smaller are the optimal explicit incentives.

- cf. "effort substitution problem" in multi-task moral hazard model of Holmstrom and Milgrom (1991)—in both cases, A's ability to privately substitute efforts across tasks or states (values of u) makes lower-powered incentives optimal.

Interaction btw. implicit and distortionary explicit incentives (BGM (94))

- Contract: (s, b, β), where implicit bonus b is to be paid iff $y = 1$ and explicit bonus β is paid iff $p = 1$

- If A trusts firm to pay b when $y = 1$, A chooses e s.t. $b + u\beta = c'(e)$.

Questions: i) Under what conditions is implicit bonus b self-enforcing? **ii)** What are optimal b and β? **iii)** Are implicit and explicit incentives substitutes or complements for firm? i.e. do optimal b and β vary with changes in environment, e.g. $var(u)$, in opposite directions (substitutes) or same direction (complements)?

- Again, focus on cooperation supported by trigger strategies.

- Now, what happens if firm reneges on implicit component of the contract depends on whether $V(\beta^*) > 0$ or $V(\beta^*) \leq 0$, and which of these applies depends on $var(u)$ and on w_a.
  - If $V(\beta^*) > 0$, then if firm reneges, parties revert to the optimal pure explicit contract (s^*, β^*).
  - If $V(\beta^*) \leq 0$, then if firm reneges, parties revert to no further interaction.

i) Under what conditions is implicit bonus b self-enforcing? Define $V(b, \beta) =$ firm's expected per-pd. profit in cooperative eqm., given that s is set so A's IR constraint binds.

- **Case 1:** If $V(\beta^*) > 0$, then NRC is $V(b, \beta) - V(\beta^*) \geq rb$.
- **Case 2:** If $V(\beta^*) \leq 0$, then NRC is $V(b, \beta) \geq rb$.

ii) What are optimal b and β? Denote these by (b^{**}, β^{**}).

Proposition: *In both Cases 1 and 2, for any b, the optimal $\beta^{**}(b)$ satisfies*

$$\beta^{**}(b) = \frac{(1-b)}{1 + var(u)} = (1-b)\beta^*, \qquad (*)$$

where $\beta^* = 1/(1 + var(u))$ is optimal in a pure explicit contract. If $b = 1$ is self-enforcing, the first-best is achieved. If $b = 1$ is not self-enforcing, then given $\beta = \beta^{**}(b)$, the optimal b is the largest positive value satisfying NRC if such a value exists, otherwise is 0.

NB: b and β are substitutes w.r.t. A's effort: $b + u\beta = c'(e)$.

iii) Are b and β substitutes or complements for firm as $var(u)$ changes?

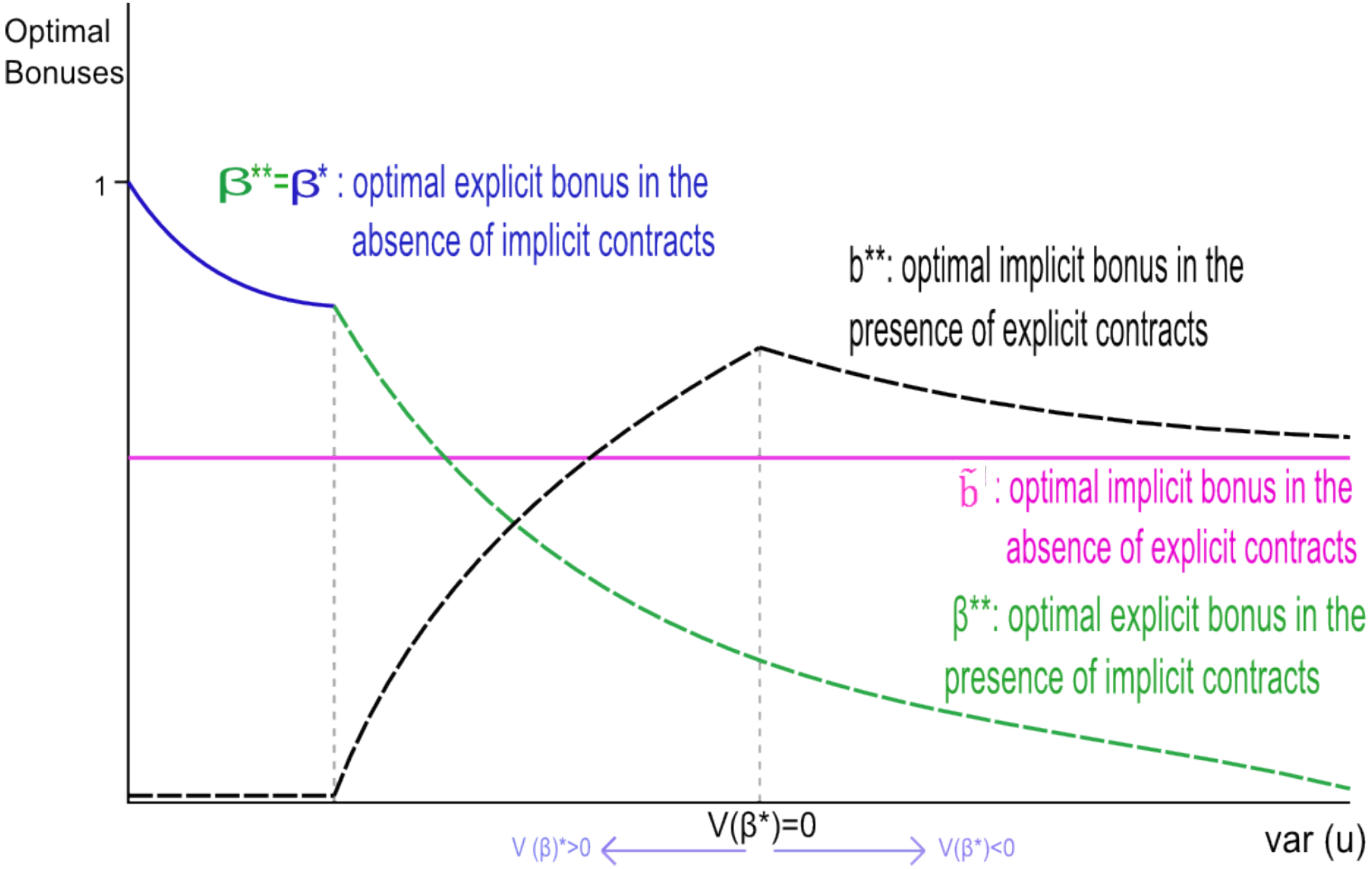
Case 1: $var(u)$ small, so $V(\beta^*) > 0$ (explicit incentives alone are profitable)

Firm chooses (b, β) to solve max $V(b, \beta)$ subj. to $V(b, \beta) - V(\beta^*) \geq rb$ (NRC).

Prop.: *If $V(\beta^*) > 0$, then as $var(u) \downarrow$, $b^{**} \downarrow$ and $\beta^{**} \uparrow$, so implicit and explicit incentives are substitutes.*

- When $var(u)$ is sufficiently low, then no implicit contract is feasible ($b^{**} = 0$), because firm's fallback position after reneging—reverting to a pure explicit contract—is too attractive. In this region, firm's maximized payoff rises as $var(u)$ falls.

- When $var(u)$ is large enough that $b^{**} > 0$, then as $var(u)$ falls, both $V(\beta^*)$ and $V(b, \beta^{**}(b))$ rise, but $V(\beta^*)$ rises faster than $V(b, \beta^{**}(b))$, so **NRC becomes tighter**. Hence b must be lowered, and $(*)$ implies β must be raised. Here, firm's maximized payoff falls as $var(u)$ falls.


Optimal
Bonuses
1
β**=β* : optimal explicit bonus in the
absence of implicit contracts
b**: optimal implicit bonus in the
presence of explicit contracts
b̃ : optimal implicit bonus in the
absence of explicit contracts
β**: optimal explicit bonus in the
presence of implicit contracts
V(β*)=0
V (β)*>0
V(β*)<0
var (u)

Case 2: $var(u)$ large, so $V(\beta^*) \leq 0$ (explicit incentives alone not profitable)

Firm chooses (b, β) to solve max $V(b, \beta)$ subj. to $V(b, \beta) \geq rb$ (NRC).

Prop.: *If* $V(\beta^*) \leq 0$, *then as* $var(u) \downarrow$, $b^{**} \uparrow$, *and* β^{**} *may* $\uparrow$ *or* $\downarrow$.

- In Case 2, as $var(u)$ falls, $V(b, \beta) \uparrow$ for any (b, β), so **NRC becomes looser**. Hence b can be raised. But a reduction in $var(u)$ also raises β^*, so $(*)$ shows there are conflicting effects on β^{**}.

- Since as $var(u)$ falls, b^{**} rises, and NRC binds, firm's maximized payoff also rises.

- When both $b^{**} \uparrow$ and $\beta^{**} \uparrow$ as $var(u) \downarrow$, then implicit and explicit incentives are *complements* – a more surprising result.

- This complementarity may help explain empirical findings that subjective bonus plans (implicit) and group piece rates (explicit) are more effective together than singly (Holmstrom-Milgrom, *AER*, 94; Ichniowski et al, *AER*, 97).

Discussion: Formally similiar interactions between relational contracts (informal arrangements) and formal contracts have been examined in other contexts:

- *Development:* publicly-provided social insurance can affect performance of informal family insurance arrangements (Di Tella and MacCulloch, *EJ*, 2002)

- *Macro:* public insurance provided through redistributive taxation may crowd out private insurance arrangements that are subject to limited enforcement (Krueger and Perri, 2011)

- *International:* "globalization", by expanding (fallback) market alterntives, can potentially impede risk-sharing in long-term employment relationships (McLaren and Newman, 2002; Spagnolo, 2002)

- *Experiments:* interaction between informal and formal arrangements (Lazzarini, Miller, and Zenger, *JLEO*, 2004)

More general analysis of relational contracts: Levin, *AER*, 2003

- A chooses effort privately in each period (HA); may have i.i.d. (over time) private information about cost of effort (HI).
- P and A risk-neutral and have time-invariant outside options; infinitely-repeated relationship with exponential discounting
- More general, time-invariant, distribution of output y_t conditional on effort e_t
- General non-stationary forms of relational contracts allowed
- Study cooperation supported by trigger strategies: trade is terminated if promised payments are ever not made.

Prop.: *If output is observable by both P and A (though not verifiable), then among optimal relational contracts (those that maximize expected surplus), there are ones that are* **stationary** *along the eqm path, i.e. the compensation schedule and effort rule are independent of time and history.*

More general analysis of relational contracts: Levin, *AER*, 2003

Prop.: *If output is observable by both P and A (though not verifiable), then among optimal relational contracts (those that maximize expected total surplus), there are ones that are* **stationary**, *i.e. the effort rule and the compensation schedule are independent of time and history.*

- There are two ways of giving incentives to A: i) discretionary transfers and ii) future conduct of relationship; for risk-neutral A, perfect substitutes in their effect.

- Changing transfer to punish A benefits P, while adjusting continuation behavior could punish both P and A. Yet with output observable by both P and A, it is never necessary to punish both P and A by destroying future surplus.

- Hence, all incentives are optimally provided through discretionary transfers to A. Moreover, A's risk-neutrality implies there's no gain to spreading his monetary rewards for a given success over multiple periods.

- NB: Availability of monetary transfers distinguishes relational contract analysis from repeated-game analysis, and is key to stationarity result.

More general analysis of relational contracts: Levin, *AER*, 2003

Prop. *Optimal stationary contracts maximize expected total surplus subj. to incentive-compatibility constraint for A and subj. to* **dynamic enforcement constraint**

$$max.payment - min.payment \leq \frac{\delta}{1-\delta}(s - \overline{s}), \qquad (DE)$$

where s is expected total surplus and $\overline{s}$ is the sum of P's and A's outside options.

- (DE) captures the no-reneging/no-quitting constraints for both P and A

Prop. *Optimal stationary contracts under hidden action (HA) alone display "information compression": A receives only 2 possible payments, high (low) if output is above (below) a threshold.*

- For a risk-neutral A, incentives should be as strong as possible subj. to (DE) $\Longrightarrow$ one-step payment rule optimal

More general analysis of relational contracts: Levin, *AER*, 2003

Now introduce **subjective performance measurement**: output is **privately** observed by P.

Prop. *An optimal relational contract in which P reports y_t in every period is a* **termination contract** *specifying a time-invariant effort (as long as relationship continues). If output is below a threshold, payment is low and relationship is terminated, while if output is above that threshold, payment is high and relationship continues.*

- Now it is impossible, with a stationary contract, to simultaneously provide P with incentives for truthful reporting of output and A with incentives for effort.

- So contract must call for destroying future surplus along the eqm path when P reports that output is low.

Dynamics within relational contracts

Recent literature studies the **dynamics within relational contracts** in environments where, because of additional sources of asymmetric information, optimal relational contracts are not stationary.

- e.g. Li and Matouschek, *AER*, 2013: If firm privately informed about opportunity cost of paying bonus, optimal relational contract generates periodic conflicts during which effort and expected profit fall, but relationship never terminates.
  - cf. Yared, *JET*, 2010: uses a similar model to analyze dynamics of relationship between an aggressive (A) and a peaceful (P) country, where A demands transfers, P is privately informed about cost of paying, and A can respond to non-payment by going to war.

- e.g. Chassang, *AER*, 2010: Analyzes how players "learn to cooperate" over the course of a relational contract, when at the start, there is asymmetric information about which actions the better-informed player "should be taking". Initially, while learning occurs, there may be inefficient termination, but if not, learning eventually stops, and parties adopt a "routine" for cooperating.

Building effective relational contracts

- Relational contracts allow parties to make use of information (e.g. A's true contribution to firm value) that is observable but not verifiable by a court.
- Non-verifiability of information used requires agreements to be self-enforcing, i.e. **credible**.
- But relational contracts face another challenge as well: **clarity** (Gibbons and Henderson, *Org. Sci.*, 2012)
- Not only must promises to cooperate be credible, but it must also be clear to both parties which actions constitute cooperation in which settings.
- In an unpredictable environment, how to achieve such clarity?
- In unpredictable environments, do agreements based on general **principles** perform better than those based on specific **rules**?
- Gibbons et al (WP, 2020) investigate this hypothesis in a lab experiment.

Two different types of implicit incentives (= two different models of incentives deriving from a concern with "reputation"):

1. **Relational contracts (=self-enforcing contracts)**

2. **Learning about A's intrinsic characteristics**

2. Learning about agent's intrinsic characteristics: can be studied using *signaling (hidden info)* or *signal-jamming (hidden action)*

- In both signaling and signal-jamming, P *learns* from A's current performance about A's intrinsic characteristics, which are relevant to A's future performance; A has implicit incentives to change current performance to influence P's *beliefs* about A's type and hence P's *future decisions*.

- This mechanism can operate even in *finite-horizon* settings, e.g. 2 periods.

- This type of implicit incentive can be adjusted only **indirectly**, by altering the information structure or the frequency of interaction.
  - Contrast the adjustable implicit bonus in relational contracts.

2a. Learning about A's intrinsic characteristics: Signaling (hidden info)

Example: Spence (1973) model of education as a signal in labor market

- A's ability – privately observed by A
- A's effort (e.g. educational, self-promotion) – public signal
- market pays wages based on estimate of ability
- If MC of effort is lower for higher-ability agents, agents have incentives to use effort to signal higher ability and hence generate higher wage offers.

2b. Learning about A's intrinsic characteristics: Signal-jamming (hidden action)

Example: Holmstrom (1999) model of managerial career concerns

- A's ability – observed by no one
- A's effort – privately-observed by A
- A's output – public signal; depends on ability and effort
- Labor market pays wages based on its estimate of ability
- Incentives for effort derive from possibility of influencing, at the margin, market's beliefs about ability, since market *can't distinguish* between effects of ability and effort on output.

2c: Many models of implicit incentives stemming from learning are intermediate between signaling and signal-jamming

- With hidden information, A might take an unobservable action (depending on type) that generates a noisy signal. Call this "noisy signaling".

- Noisy signaling models are particularly useful for studying how reputational incentives evolve over time, i.e. in relationships lasting more than 2 periods.

Additional examples of implicit incentives stemming from learning:

- ratchet effect in regulation (firm's type is its intrinisic efficiency)
  - e.g. Laffont and Tirole (1988) – signaling; Meyer and Vickers (1997) – signal-jamming

- incentives for incumbent politicians seeking re-election (politican's type can be ability or preferences/corruptibility)
  - e.g. Coate and Morris (1995) – noisy signaling

Additional examples of implicit incentives stemming from learning:

- incentives for influence activities in firms (employee's type can be own quality or quality of project, division, etc.)
  - e.g. Meyer, Milgrom, and Roberts (1992) – signal-jamming

- incentives for prosocial behavior, e.g. contributions to charity or public goods (agent's type is weights he attaches to prosocial actions and to money)
  - e.g. Benabou and Tirole (2006) – noisy signaling (they also study interaction btw. implicit and explicit incentives)

- incentives for firms to provide high-quality products (firm's type is its ability to provide high quality)
  - e.g. Tadelis (2002) and Mailath and Samuelson (2001) – noisy signaling

- incentives for experts to provide good advice (expert's type can be ability (=quality of information) or how aligned preferences are with decision-maker's)
  - e.g. Scharfstein and Stein (1990), Ottaviani and Sorensen (2006), Prat (2005) – signal-jamming (expert does not know own ability, and expert's hidden action is the choice of how to map his information into advice)