CORPORATE SOCIAL RESPONSIBILITY IN THE GLOBAL VALUE CHAIN: A BARGAINING PERSPECTIVE
- APPENDIX -

Micro-foundations of the Value-Price-Cost Model

Several studies have modeled bargaining power in supply chains (although not explicitly considering CSR or sustainability issues). For example, Dukes et al. (2006) apply the bargaining theory to show that manufacturers reap part of the cost savings of their retailers through bargaining, while Erteg & Griffin (2002) study non-cooperative pricing decisions of buyers and suppliers. The following micro-foundations for our model build on Dukes et al. (2006) with a few changes to reflect our setup.[1] Suppose that the MNE faces demand $D(r) = a - r$, where r is the price it chooses in its output market, and that the supplier faces a marginal cost of $c > 0$ (in both cases assuming that the supplier behaves responsibly). We assume that an agreement between the two firms consists of a quantity to be produced and a price to be paid for this quantity by the MNE to the supplier. If the negotiated unit price paid by the MNE to the supplier is denoted by p, the MNE will earn $D(r)(r - p)$ and the supplier will earn $D(r)(p - c)$, for a total (coalitional) profit of $D(r)(r - c) = (a - r)(r - c)$.

As described in the paper, we apply the Nash Bargaining Solution (NBS) to arrive at a prediction. Rubinstein (1982) demonstrated how the NBS is consistent with the equilibrium of a non-cooperative game where the players make alternating offers. One of the axioms of the Nash Bargaining Solution (NBS) is that it is *efficient*, suggesting that any agreement between the two firms should stipulate the output—implying the output price—that maximizes their joint payoff. If the agreement does not, they will have an incentive to renegotiate and find a

[1] In particular, unlike Dukes et al. (2006), we assume away the marginal cost of the MNE (corresponding to the retailer in their model) and instead ascribe a marginal cost to the supplier (corresponding to the manufacturer in their model).

different solution where they are both better off. The jointly optimal output price is given by $r = (a - c)/2$, implying output of $D(r) = (a + c)/2$ and a coalitional payoff of $(a + c)^2/4$. Relating these calculations to our reduced-form model, the supplier then incurs total costs $C \equiv c(a - c)/2$, and the MNE will earn revenue of $V \equiv (a^2 - c^2)/4$, with the total value created being equal to $V - C \equiv (a - c)^2/4$ as identified above. If both firms have outside options of 0, the NBS of the unit price paid by the MNE will be $(a + c)(1 - \lambda)/2 + c\lambda$, resulting in a total payment from the MNE to the supplier of $P \equiv pD(r) = (a^2 - c^2)(1 - \lambda)/4 + (a - c)c\lambda/2$. We can see from these functions that, if we model the Value-Price-Cost framework using demand and cost curves, V and P are increasing in demand and decreasing in marginal cost, while C is increasing in marginal cost and in demand.

Propositions including mathematical conditions

Proposition 1*: After the supplier is detected by the outside world to have acted irresponsibly, it opportunistically renegotiates a higher price from the MNE, which suffers a hold-up problem, when:*

a. *The reputational benefit to the MNE from terminating the relationship is low, that is* $F < V - C - O_M - O_S - W$.
b. *The MNE has limited bargaining power and experiences large reputational damage, relative to the supplier, that is* $\lambda < B_F/(B_F + B_S)$.

Proposition 2*: Social branding investments that increase the MNE's reputational damage from working with an irresponsible supplier improves MNE performance by making the threat of terminating the relationship with an irresponsible supplier credible when:*

a. *The value of the relationship, including the cost of correcting the problem, is positive but small compared to the effect of the social branding investments, that is* $0 < [(V - C - O_M - O_S) - W] - F < Z$;
b. *The supplier is powerful and resilient to reputational damage, but has a low outside option, that is* $(2B_S + O_S)/(1 - \lambda) < Z$.

Proposition 3.1*: After detecting an irresponsible behavior, the MNE chooses to announce it publicly when:*

- *The overlap between internal and external monitoring is high, the reputational damages are small, the reputational discounts from being proactive and from terminating the relationship are large, and the bargaining strength of the supplier is high, that is* $(1-\lambda)F > (1-\alpha)(B_S + B_M - R)$.

Proposition 3.2*: Credible commitment to announcing the problem increases MNE performance by influencing supplier incentives to behave responsibly when:*

- *The cost savings from irresponsible action is moderate, that is* $\Delta_Q < \Delta < \Delta_A$ *(see below for definitions of these thresholds).*

Proposition 4.1*: Monitoring an irresponsible supplier influences the performance of the MNE negatively (even excluding the operational costs of the monitoring itself) when:*

- *The overlap between the internal and the external monitoring is intermediate, that is* $\frac{O_M+R}{B_M-F+(F+W+\Delta)\lambda} < (1-\alpha) < \frac{(1-\lambda)F}{B_S+B_M-R}$.

Proposition 4.2*: Credible commitment to monitoring the supplier can influence the performance of the MNE negatively, by incentivizing the supplier to be irresponsible, when:*

- *The MNE is expected to work privately with the supplier to fix the problem if discovered through monitoring, that is* $(1-\lambda)F < (1-\alpha)(B_S + B_M - R)$.

Proof of proposition 1

First, we characterize the scenario in which the MNE will keep the supplier when revealed to be irresponsible. When the two firms work to fix the problem, they earn a coalitional payoff of $V - C - W - B_M - B_S$. On the other hand, when they terminate the relationship, the MNE earns $O_M - B_F$ and the supplier earns $O_S - B_S$. The coalitional payoff is higher than the sum of the outside options if $V - C - O_M - O_S > W + B_M + B_S - B_S - B_F = W + F$, which is the inequality provided in the paper and assumed in the following to be fulfilled.

Without loss of generality, we assume that W falls on the supplier side. If the two firms could commit to a price in advance, they would commit to sharing the total surplus according to their bargaining strength in any scenario that could arise, such that incentives are aligned. This means that the MNE would be willing to accept in advance (when drawing up the contract) a price in the case of detection such that $V - P - B_M > O_M \Leftrightarrow P < V - B_M - O_M$ and the supplier a price such that $P - W - B_S - C > O_S \Leftrightarrow P > C + O_S + W + B_S$. These reservation prices are shown in the third column in Figure 3 in the paper. The NBS is $P_D^* = \lambda(C + O_S + W + B_S) + (1 - \lambda)(V - B_M - O_M)$. However, once detection takes place, the supplier can opportunistically renegotiate and hence any initial renegotiations will be irrelevant to the outcome. At that point in time, the MNE earns $V - P - B_M$ by working with the supplier and $O_M - B_F$ by terminating the relationship, while the supplier earns $P - B_S - W$ by working with the MNE and $O_S - B_S$ if the relationship is terminated. Accordingly, the MNE will accept a price of $P < V - B_M - O_M + B_F = V - O_M - F$, which is higher than the price it would have commited to in advance (by the magnitude $B_M - F = B_F$). Similarly, the supplier will accept a price of $P > C + O_S + W$, which is lower than the price it would have committed to. These reservation prices are shown in the fourth column in the figure. The NBS is $P_D = \lambda(C + O_S + W) + (1 - \lambda)(V - O_M - F)$. We can then solve the inequality $P_D > P_D^*$ to arrive at $\lambda < B_F/(B_F + B_S)$, which is condition (b) in Proposition 1.

We now examine when this leads to a conflict of interest where the behavior of the supplier damages the MNE. Let π_{MD} (π_{SD}) be the profit of the MNE (supplier) when the supplier is irresponsible and detected, π_{MN} (π_{SN}) be the profit of the MNE (supplier) when the supplier is irresponsible and not detected, and π_{MR} (π_{SR}) be the profit of the MNE (supplier) when the supplier is responsible, as shown in Figure 3. These profit functions are given by:

$$\pi_{MD} = (V - O_M - B_M) - P_D$$

$$\pi_{SD} = P_D - (C + O_S + W + B_S)$$

$$\pi_{MN} = (V - O_M) - P_N$$

$$\pi_{SN} = P_N - (C + O_S - \Delta)$$

$$\pi_{MR} = (V - O_M) - P_R$$

$$\pi_{SR} = P_R - (C + O_S).$$

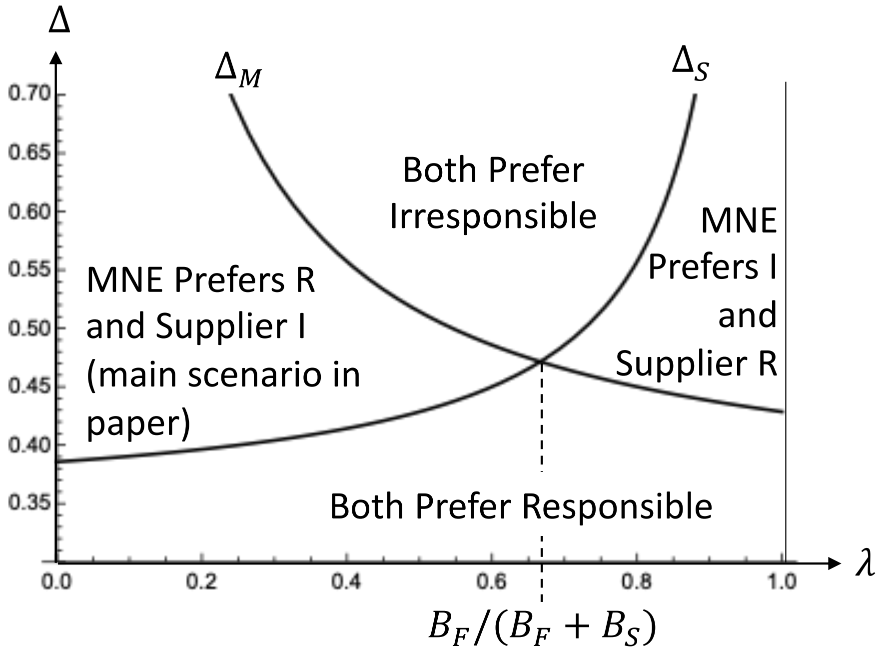
Where $P_N = \lambda(C + O_S - \Delta) + (1 - \lambda)(V - O_M)$ and $P_R = \lambda(C + O_S) + (1 - \lambda)(V - O_M)$. Note that P_N is decreasing in Δ and hence, when not detected, the MNE reaps (proportionally to its bargaining power) a part of the unobserved cost savings from the irresponsible action of the supplier.[2] The MNE will prefer the supplier to be responsible if $\pi_{MR} > x\pi_{MD} + (1 - x)\pi_{MN}$, and the supplier will prefer to be irresponsible if $\pi_{SR} > x\pi_{SD} + (1 - x)\pi_{SN}$. Denote by Δ_M the threshold value of Δ that solves the first inequality, and by Δ_S the threshold value solving the latter. Then, it can be shown that $\Delta_M > \Delta_S$ as long as at $\lambda < B_F/(B_F + B_S)$, which is the exact same condition derived for the hold-up problem to occur. In other words, when the hold-up problem occurs, there is a range of moderate cost savings, $\Delta_S < \Delta < \Delta_M$ in which the supplier prefers to be irresponsible while the MNE prefers it to be responsible. This range is wider (and hence the conflict more likely to occur) when F, λ, and B_S are lower, and x and B_M are higher.

We have still not explored the possibility that $V - C - O_M - O_S < W + F$, in which case there is no quasi-surplus left to bargain over once the supplier's irresponsible behavior has been detected. In that case, as described above, the firms will terminate the relationship, and the MNE will get $O_M - B_F$ while the supplier gets $O_S - B_S$. However, because both firms suffer reputational damage outside of the relationship, and this damage may be higher for one firm than the other, they may not have equally much to lose. In particular, it can be shown that the

[2] Formally, we assume that the MNE has a belief about the supplier's behavior (e.g. based on its knowledge about the supplier and the environment in which it is operating), which is fulfilled in equilibrium. Because it is based on a belief, the MNE cannot prove, correct, or publicly announce the irresponsible actions of the supplier. On the flip side, this situation allows for plausible deniability, as we shall see later.

MNE has more to lose from the supplier's irresponsible behavior than the supplier does if $\lambda < B_F/(B_F + B_S)$. This is the same condition as before, which is not coincidental since it is the same mechanism—reputational damage outside the relationship—that creates both the hold-up problem and the potential conflict of interest when the relationship is terminated.

As mentioned in the discussion section in the paper, this conflict of interest is only one among many possibilities. In the figure below, we plot Δ_S and Δ_M as a function of λ in order to show how the preferences of the players may converge or diverge. As indicated in this figure (shown for $V - C - O_M - O_S > W + F$, $W = B_M = 0.5$, $x = F = 0.3$, $B_S = 0.1$), all four combinations of preferences are possible.



Proof of Proposition 2

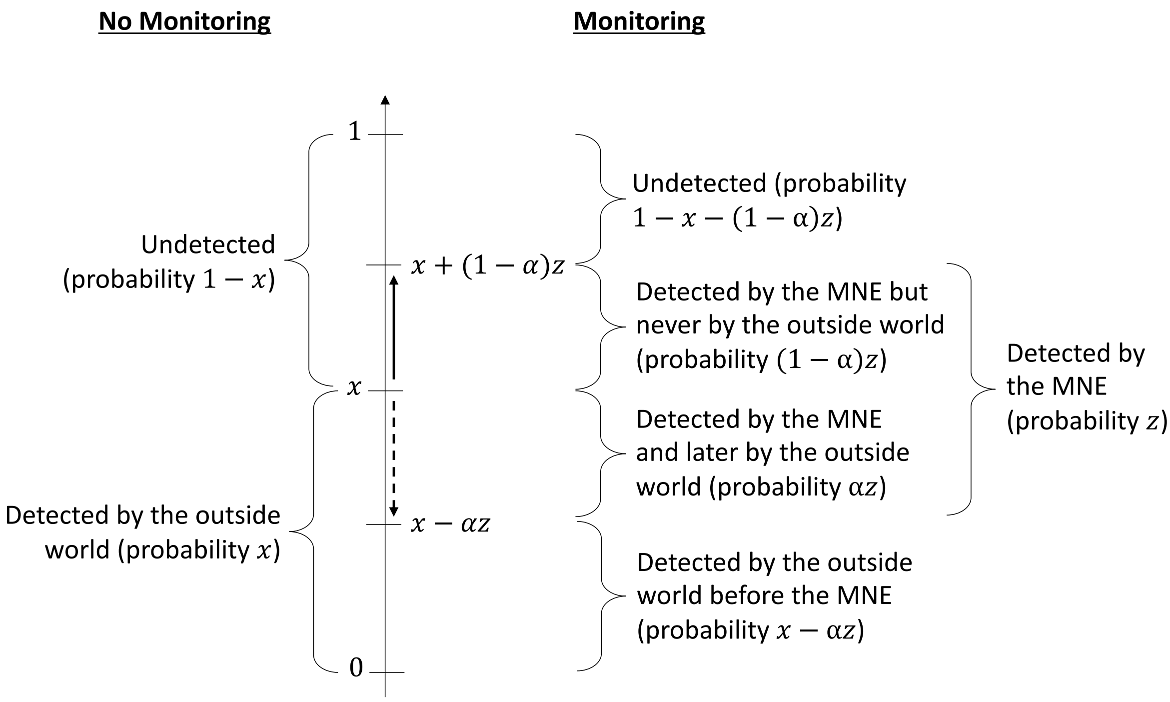
As long as $\Delta_S < \Delta < \Delta_M$, the MNE can improve its performance by inducing a change in the supplier's behavior from irresponsible to responsible. Also, if the MNE is successful in doing so, any change in the reputational damage parameters will have no (negative) effect in equilibrium (because the supplier will be responsible). Hence, the social branding effort of the MNE must have two consequences: (1) it must eliminate any quasi-surplus in the case of detection, such that the MNE can credibly commit to terminate the relationship, and (2) the threat of relationship termination must create sufficient incentives for the supplier to change its behavior. We examine each condition in turn.

First, suppose that the MNE can increase the reputational damage from working with an irresponsible supplier from B_M to $B_M + Z$, while leaving B_F unchanged. This implies that F also rises by a similar amount (Z), since $F = B_M - B_F$. For the change to have an effect, it is necessary that $V - C - O_M - O_S > W + F$ (the relationship is not terminated in the baseline scenario) and that $V - C - O_M - O_S < W + F + Z$ (the relationship is terminated if the social branding is considered). Clearly, this requires that neither the value of the relationship ($V - C - O_M - O_S$), the cost of working to correct the problem (W), or the original reputational discount (F) can be exceedingly high or low, because then one of these inequalities would not hold.

Second, the irresponsible supplier now can expect to earn $x\pi_{SD} + (1-x)\pi_{SN}$, where $\pi_{SD} = O_M - B_S$. It will choose to be responsible if π_{SR} is higher than that, and the resulting inequality can be solved to find a threshold value of cost savings which we denote Δ_{SF}. Within the range $[\Delta_S, \Delta_{SF}]$, the supplier will be irresponsible without the social branding, but responsible with it. This range exists only if $\Delta_S < \Delta_{SF}$, which can be solved for $V - C - O_M - O_S > W + F + (2B_S + O_S)/(1-\lambda)$. Comparing with the two inequalities above, we can see that this does not rule out the first one, but it does rule out the second one unless $(2B_S + O_S)/(1-\lambda) < Z$. Hence, this scenario can only take place if Z is sufficiently high, and λ, B_S and O_S are sufficiently low.

Modeling monitoring probabilities

Let z be the probability that the monitoring of the MNE is successful in detecting the irresponsible behavior and $\alpha \in [0,1]$ be the likelihood that such irresponsible behavior is subsequently revealed to the world. α captures therefore the degree of "overlap between internal and external monitoring". We assume that $x + z < 1$ and $z < x$. There are four mutually exclusive outcomes: (1) With probability $x - \alpha z > 0$, the outside world discovers an irresponsible behavior before the MNE does; (2) With probability αz, the MNE discovers an irresponsible behavior, which will later be detected by the outside world, but now the MNE has an opportunity to act on it first; (3) With probability $(1 - \alpha)z$, the MNE discovers an irresponsible behavior that will never be detected by the outside world, unless the MNE does something to bring attention to it; (4) With probability $1 - x - (1 - \alpha)z$, the irresponsible behavior is undetected by the MNE as well as by the outside world. These probabilities are illustrated in the figure below.

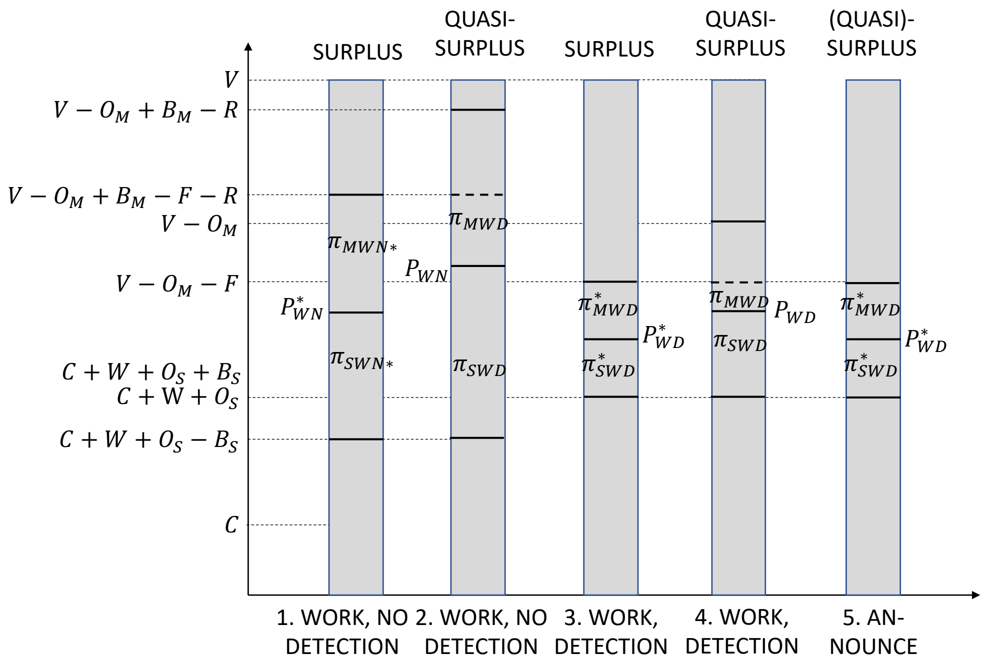


By adding together (2) and (3), as shown in the middle part of the outcome space in the figure, we have z, the probability that the MNE monitoring is successful. However, the MNE does not know whether the outside world will also discover the irresponsible behavior later on (2) or not (3)—in other words, whether the draw from the distribution is above or below x. Intuitively, this creates a potential dilemma as to the magnitude of the MNE response. If α is high, the dashed arrow will be long and the solid arrow will be short, and hence it is highly likely that the irresponsible behavior discovered by the MNE will be revealed at a later point in time. In that case, we can say that the internal monitoring of the MNE is a substitute for the external monitoring of NGOs and media. On the other hand, if α is low, the dashed arrow will be short and the solid arrow will be long, and hence the irresponsible behavior discovered by the MNE will likely not be revealed subsequently. In that case, the internal monitoring of the MNE is additive to the external monitoring. If the MNE discovers an irresponsible behavior, it updates its belief about whether the irresponsible behavior will later be detected by the outside world. Using Bayesian logic, we can see that, given that the MNE has discovered an irresponsible behavior, the likelihood that it will later be detected by the outside world is α. In other words, the MNE can infer that the draw is between $x - \alpha z$ and $x - (1 - \alpha)z$, and, accordingly, there is a perceived 'residual probability' of $\alpha z/z = \alpha$ that the outside world will detect the irresponsible behavior later on.

Proof of Proposition 3

We start by solving the four-way choice when the MNE detects a problem that the outside world has not (yet) discovered. First, the choice to monitor and ignore the problem is never going to occur in equilibrium. If the MNE chooses to ignore the problem, it puts itself in the same situation as if it had never done any monitoring, except that now it suffers d on top of the normal reputational damage if the irresponsible behavior is revealed. Hence, monitoring and ignoring the problem is strictly dominated by not monitoring, no matter the cost of monitoring. In addition, monitoring to ignore has no strategic effect on the supplier's actions and therefore serves no deterrence purpose either.

This means that the choice is between working privately to correct the problem, announcing the problem and working with the supplier, and terminating the relationship. If they terminate the relationship, the MNE immediately gets $O_M - B_M + F + R$ and the supplier gets $O_S - B_S$. If they continue the relationship and work privately to correct the problem, there are two possible outcomes: one in which the problem is detected by the outside world afterwards, and one in which it is never detected. Suppose first that the two firms could specify in a contract the price under each of these possible contingencies. The resulting prices are shown in scenarios (1) and (3) in the figure below.



In scenario 1, the irresponsible behavior is never revealed to the outside world. Hence, the MNE earns $V - P$, which must be better than what it will get from terminating the relationship immediately after monitoring, $O_M - B_M + F + R$, so it would accept a price in advance where $P < V - O_M + B_M - F - R$. The supplier earns $P - V - C$, and it would earn $O_S - B_S$ if the relationship is terminated, so it will accept in advance a price $P > C + W + O_S - B_S$. The NBS would be P^*_{WN}.

Like in the non-monitoring scenario, the problem is that it is not subgame perfect to stick to such a price, and hence the supplier can opportunistically renegotiate the price after successfully avoiding detection by the outside world (see scenario 2). If the relationship is terminated at that time, the irresponsible behavior is revealed, and reputational damages of $B_M - R$ and B_S are incurred—but the MNE can no longer get F once it has chosen to continue to work with the supplier. In other words, it will look worse if the MNE discovers the problem but waits to terminate the relationship until it is later revealed to the world, compared to terminating the relationship immediately upon detection. By continuing to work together, the MNE still earns $V - P$, but now its outside opportunity has deteriorated to $O_M - B_M + R$ because it has chosen not to terminate the relationship immediately upon discovering the problem, and so misses out on F. This means its willingness to pay increases to $P < V - O_M + B_M - R$. The supplier will still accept any price $P > C + W + O_S - B_S$, and therefore the price increases to P_{WN}. The two firms are now bargaining over quasi-surplus and the supplier is holding up the MNE to get a larger share of the surplus.

A similar situation arises in case the problem is detected by the outside world afterwards. Assuming again that an enforceable price can be contracted upon (see scenario 3), the MNE would then earn $V - (B_M - R) - P$, and that needs to be better than its outside opportunity if it the relationship is terminated immediately, $O_M - (B_M - F - R)$. Hence, the MNE is willing to agree in advance to a price satisfying $V - (B_M - R) - P > O_M - (B_M - F - R)$, implying that $P < V - O_M - F$. The supplier earns $P - C - W - B_S$ in the future and can leave immediately and get $O_S - B_S$, so it is willing to accept in advance a price $P > C + W + O_S$. The NBS would be P^*_{WD}.

However, when the situation arises and the supplier's irresponsible behavior is detected, it can opportunistically renegotiate (see scenario 4). At the time of this renegotiation, the MNE's outside opportunity has changed to $O_M - (B_M - R)$, because it can no longer look righteous and get F by terminating the relationship. Hence, the MNE is willing to agree at this point in time to a price satisfying $V-(B_M-R)-P>O_M-(B_M-R)$, implying that $P<V-O_M$. The supplier can still get O_S-B_S by leaving and will earn $P-C-W-B_S$ by staying, so it is willing to accept a price of $P>C+W+O_S$. This means the price will rise to P_{WD}. Intuitively, the MNE is unable to make the supplier bear some of the incremental reputational damage from detection, which therefore becomes quasi-surplus that is up for bargaining.

In both scenarios, the quasi-surplus includes the actual surplus but with the addition of F. By choosing to work privately with the supplier after monitoring, the MNE puts itself in a tenuous position: Since it can no longer choose its outside option to include the reputational discount F from terminating the relationship, it is vulnerable to hold-up by the supplier by that amount.

Finally, the MNE can choose to announce the irresponsible behavior, but still work with the supplier to fix it (see scenario 5). In that case, the MNE retains the threat of terminating the relationship and reaping F if negotiations fail, so its outside option increases to $O_M-(B_M-F-R)$ while it earns $V-(B_M-R)-P$ by working with the supplier. Therefore, its willingness to pay becomes $P<V-O_M-F$. The supplier can still get O_S-B_S by leaving and will earn $P-C-W-B_S$ by staying, so it is willing to accept a price of $P>C+W+O_S$. As the figure shows, this enables the MNE to achieve P^*_{WD}, which corresponds to its share of the true surplus upon detection, which it could not obtain if it worked privately with the supplier because of the hold-up problem. Hence, announcing the existence of an irresponsible behavior protects the MNE against the hold-up problem.

We now assume that the relationship is sufficiently valuable so that the MNE will not terminate it, leaving it with the choice between announcing the problem and working privately to correct it. Announcing the problem results in $\pi_{MA}=V-B_M+R-P^*_{WD}$, whereas keeping it private results in $\pi_{MWD} = V - B_M + R - P_{WD}$ if subsequently detected by the outside world and $\pi_{MWN} = V - P_{WN}$ if not. It will be better to announce the problem if $\pi_{MA} > \alpha\pi_{MWD} + (1-\alpha)\pi_{MWN}$, which holds for $(1-\lambda)F > (1-\alpha)(B_S + B_M - R)$, as stated in the condition for proposition 3.1. From this inequality, we can see that announcing is more likely to be attractive when λ, B_S, and B_M are low, and F, α, and R are high.

To prove proposition 3.2, we have to look at the incentives of the supplier to behave responsibly on the expectation of each of the two outcomes. Suppose the supplier expects the MNE to announce the irresponsible behavior if monitoring is successful. Then, the irresponsible supplier earns $\pi_{SA} = P_{WD}^* - C - W - B_S$ if detected by the MNE, π_{SD} if undetected by the MNE but later detected by the outside world, and π_{SN} if never detected by anyone. Hence, it is better to be responsible if $\pi_{SR} > z\pi_{SA} + (x - \alpha z)\pi_{SD} + (1 - x - (1-\alpha)z)\pi_{SN}$. Let Δ_A denote the threshold value determining whether the supplier will be responsible or not, based on the solution to this inequality.

Suppose instead that the supplier expects the MNE to work privately upon successful monitoring. Then, the irresponsible supplier earns $\pi_{SWD} = P_{WD} - C - W - B_S$ if detected first by the MNE and subsequently by the outside world, $\pi_{SWN} = P_{WN} - C - W$ if detected by the MNE but never detected by the outside world, π_{SD} if undetected by the MNE but later detected by the outside world, and π_{SN} if never detected by anyone. Hence, it is better to be responsible if $\pi_{SR} > \alpha z\pi_{SWD} + (1-\alpha)z\pi_{SWN} + (x - \alpha z)\pi_{SD} + (1 - x - (1-\alpha)z)\pi_{SN}$. Let Δ_Q denote the threshold value determining whether the supplier will be responsible or not, based on the solution to this inequality.

It can be shown that $\Delta_A > \Delta_Q$, suggesting that there is a range of intermediate cost savings in which the supplier will be responsible only if it expects the MNE to announce the problem. The width of the range is $\Delta_A - \Delta_Q = z((1-\alpha)(B_S + B_M - R) + F)/(1 - x - (1-\alpha)z)$. Hence, it is more likely to happen (the range is wider) if the denominator—which is the likelihood of the supplier completely avoiding detection—is low. It is also more likely if B_S, $B_M - R$, and F are high.

Proof of Proposition 4

To prove proposition 4.1, we first assume that the supplier is irresponsible, and the MNE believes that it is so but has no proof or concrete observations of the irresponsible behavior. The question then is how the MNE performance is affected by monitoring. We assume here that the relationship is sufficiently valuable that the MNE will not terminate it, whether discovered by the MNE itself or by the outside world. If the MNE does not monitor, it earns, as before, $x\pi_{MD} + (1-x)\pi_{MN}$. If it monitors, there are three possibilities in the first instance:

1) With probability $1 - x - (1-\alpha)z$, the supplier's irresponsible behavior is never detected by anyone and the MNE earns π_{MN}.
2) With probability $x - \alpha z$, the supplier's irresponsible behavior is not detected by the MNE, but later by the outside world, and the MNE earns π_{MD}.
3) With probability z, the supplier's irresponsible behavior is detected by the MNE, which can choose between announcing the problem and working privatey to fix it, as described above.

Assume first that $(1-\lambda)F > (1-\alpha)(B_S + B_M - R)$ (as implied by condition (a) of proposition 4.1) such that announcing the problem is subgame perfect for the MNE in scenario 3. In that case, it earns π_{MA}. This means that monitoring comes with expected profits of $(1 - x - (1-\alpha)z)\pi_{MN} + (x - \alpha z)\pi_{MD} + z\pi_{MA}$. The difference between profits with and without monitoring (henceforth, the benefit of monitoring) can be calculated to be:

$$z(O_M + R - (1-\alpha)(B_M - F + (F + W + \Delta)\lambda)).$$

This can be either positive or negative. To see that it can be positive, suppose that the overlap between internal and external monitoring is high ($\alpha = 1$). In that case, the MNE will always announce because $(1-\lambda)F > 0$ and the benefit of monitoring reduces to $z(O_M + R)$, which is always positive. To see that it can be negative, suppose now that the overlap between internal and external monitoring is intermediate ($\alpha = 1/2$), the MNE has weak bargaining strength ($\lambda = 1/3$) and gets no reputational benefit from being proactive ($R = 0$), the supplier suffers no brand damage ($B_S = 0$), and the MNE has a lousy outside option ($O_M = 0$). In that case, the MNE will announce if $F > \frac{3}{4}B_M$, while monitoring has a negative benefit if $F < \frac{3}{2}B_M + W + \Delta$, which is always the case because $F < B_M$ by definition. Hence, there is a range $\frac{3}{4}B_M < F < B_M$ where the MNE will announce the problem when discovered, and because of that its performance is lower when monitoring. This proves that the two jointly sufficient conditions of proposition 4.1 can be fulfilled simultaneously. Looking at the general expression for monitoring benefits, we can see that it is more likely to be positive when α, O_M, and R are high, and when B_M, W and Δ are low.

Combining the conditions that $(1-\lambda)F > (1-\alpha)(B_S + B_M - R)$ and $z\big(O_M + R - (1-\alpha)(B_M - F + (F + W + \Delta)\lambda)\big) < 0$ and rearranging gives us the following expression, capturing the condition of intermediate overlap: $\frac{O_M+R}{B_M-F+(F+W+\Delta)\lambda} < (1-\alpha) < \frac{(1-\lambda)F}{B_S+B_M-R}$.

To prove proposition 4.2, we need to look at the supplier's incentives when the MNE is committed to monitoring or to not monitoring. When the MNE does not monitor, we have already seen that the supplier will be irresponsible if $\Delta > \Delta_S$. Suppose that the MNE does monitor and that it will work privately with the supplier if the irresponsible behavior is detected—this requires, as we have seen above, that $(1-\lambda)F < (1-\alpha)(B_S + B_M - R)$ and that the relationship is sufficiently valuable (high V and low C). The first of these conditions is necessary for the proposition to hold, because if it is not fulfilled and the MNE will announce the problem if discovered, we have already seen that monitoring cannot make the irresponsible supplier better off (because $\Delta_A > \Delta_Q$). The question is whether there are parameter configurations consistent with this condition that makes the irresponsible supplier better off. When the MNE will not announce, all four scenarios come into play:

1) With probability $1 - x - (1 - \alpha)z$, the supplier's irresponsible behavior is never detected and earns π_{SN}.
2) With probability $x - \alpha z$, the supplier's irresponsible behavior is not detected by the MNE, but later detected by the outside world, and earns π_{SD}.
3) With probability αz, the supplier's irresponsible behavior is detected by the MNE, who first works privately to address it, and later detected by the outside world, resulting in profits of π_{SWD} to the supplier.
4) With probability $(1 - \alpha)z$, the supplier's irresponsible behavior is detected by the MNE, who works privately to address it, and never revealed to the outside world, resulting in profits of π_{SWN} to the supplier.

Given these probabilities, the expected profits of the irresponsible supplier when the MNE is monitoring is equal to:

$$(1 - x - (1 - \alpha)z)\pi_{SN} + (x - \alpha z)\pi_{SD} + \alpha z \pi_{SWD} + (1 - \alpha)z\pi_{SWN}.$$

Let Δ_N be the threshold value of Δ above which these profits are higher than the profits of a responsible supplier. The question then is how Δ_N and Δ_S compares. If $\Delta_N > \Delta_S$, there is a range (between these values) where the supplier will be irresponsible if not monitored but responsible otherwise, while if $\Delta_N < \Delta_S$ there is a range (between these values) where monitoring in fact encourages irresponsible action. Again, both scenarios are possible. Suppose that $\alpha = x = z = \lambda = 1/2$. In that case, it can be shown that $\Delta_N - \Delta_S = 3B_S - (B_M - R) - 4O_S + 2W$, which can be positive or negative. It is more likely to be negative (so that monitoring has an unintended consequence of making the supplier irresponsible) if the brand damage suffered by the supplier (B_S) is small compared to the brand damage suffered by the MNE ($B_M - R$), if the supplier has an attractive outside opportunity (low O_S), and if the cost of working to correct the problem is low (low W). The expression can be negative without ruling out the condition for the MNE to work privately, which is fulfilled as soon as F is low enough. This proves that the necessary condition for the proposition can be fulfilled simultaneously with other conditions that are required to be sufficient.

REFERENCES (not in the paper)

Ertek G., P. M. Griffin. 2002. Supplier- and buyer-driven channels in a two-stage supply chain. *IIE Transactions*, *34*, 691-700.

Rubinstein, A. 1985. A bargaining model with incomplete information about time preferences. *Econometrica 53*, 1151–1172.