— Online Appendix —

B Additional Proofs

B.1 Proof of Proposition 3

Proof. For the proof of this and the following proposition, it is useful to introduce the terms

$$T(y) := F(y)\,(y - \mathbb{E}[\theta - \beta|\theta \le y])\,,$$

and

$$S(y) := (1 - F(y))\,(y - \mathbb{E}[\theta - \beta|\theta \ge y])\,,$$

in the literature referred to as, respectively, backward bias and forward bias (see Alonso and Matouschek 2008). By condition (7) we have

$$T''(y) = \beta f'(y) + f(y) > 0 \quad \text{and} \quad S''(y) = -(\beta f'(y) + f(y)) < 0 \quad \text{for all } y \in [0,1]$$

Note first that - since $\beta > 0$ - we have $T(y) \ge 0$ for all $y \in [y_{min}, y_{max}]$ and $T(y) > 0$ for $y \ge 0$. The variable S may change sign. Noticing however, that $S(\hat{y}) = S(1) = 0$, strict concavity of S implies that $S(y) > 0$ for all $y \in (\hat{y}, 1)$.

Having introduced these terms, we are ready to prove Proposition 3. This proof is presented via three lemmas.

Lemma 11. *Let condition (7) be satisfied and consider $y_1, y_2 \in Y$ with $y_1 < y_2$. If $y_1, y_2 \in D^*(Y)$, then all $y \in (Y \cap (y_1, y_2))$ belong to $D^*(Y)$.*

Proof. Towards a contradiction suppose there is some $y \in Y$ such that $y \notin D^*(Y)$ and $D^*(Y) \cap [y_{min}, y] \neq \emptyset$, $D^*(Y) \cap [y, y_{max}] \neq \emptyset$. Further, let y^- be the largest element of $D^*(Y)$ strictly smaller than y and let y^+ be the smallest element of $D^*(Y)$ strictly greater than y, that is $y^- = \max\{y' \in D^*(Y) : y' < y\}$ and $y^- = \min\{y' \in D^*(Y) : y' > y\}$. Define $s := \frac{y^- + y^+}{2}$ to be the state at which the agent is indifferent between choosing action y^- and action y^+, and similarly define $r := \frac{y + y^-}{2}$ and $t := \frac{y^+ + y}{2}$ as the states in which the agent is indifferent, respectively, between choosing y^- and y and between y^+ and y.

Following Alonso and Matouschek (2008), we can write the change in the principal's expected payoff when including action y into the delegation set. The agent changes his choice of action only in states $[r, t]$. In states $[r, s]$ he switches from y^- to y, while in the remaining states $(s, t]$

he switches from y^+ to y. The change in the principal's expected payoff is thus given by

$$
\begin{aligned}
& -\int_r^t (y-\theta+\beta)^2 f(\theta)\mathrm{d}\theta + \int_r^s (y^- -\theta+\beta)^2 f(\theta)\mathrm{d}\theta + \int_s^t (y^+ -\theta+\beta)^2 f(\theta)\mathrm{d}\theta, \\
= \quad & 2(y-y^-)\underbrace{F(r)\left[r-\mathbb{E}\left[\theta-\beta|\theta\leq r\right]\right]}_{=T(r)} + 2(y^+-y)\underbrace{F(t)\left[t-\mathbb{E}\left[\theta-\beta|\theta\leq t\right]\right]}_{=T(t)} \\
& -2(y^+-y^-)\underbrace{F(s)\left[s-\mathbb{E}\left[\theta-\beta|\theta\leq s\right]\right]}_{=T(s)}.
\end{aligned}
$$

Letting $y = \lambda y^+ + (1-\lambda)y^-$ for some $\lambda \in (0,1)$ so that $y - y^- = \lambda(y^+ - y^-)$, $y^+ - y = (1-\lambda)(y^+ - y^-)$ and $s = \lambda r + (1-\lambda)t$, the payoff difference can be written as

$$2(y^+ - y^-)\left[\lambda T(r) + (1-\lambda)T(t) - T(\lambda r + (1-\lambda)t)\right].$$

From the strict convexity of T, it then follows that the payoff difference is strictly positive. A contradiction. □

Lemma 12. *The optimal delegation set satisfies* $\min D^*(Y) = \min Y$.

Proof. Consider delegation set D with $\min D(Y) > \min Y$. Letting $y = \min Y$ and $\underline{y} = \min D(\widehat{y})$, the state at which the agent is indifferent between the two actions is given by $t = (y + \underline{y})/2$. If the principal includes y in the delegation set, the agent switches from $\underline{y}$ to y in all states $\theta \leq t$. The principal's change in expected payoff when including y is hence given by

$$
\begin{aligned}
& -\int_0^t (y-\theta+\beta)^2 f(\theta)\mathrm{d}\theta + \int_0^t (\underline{y}-\theta+\beta)^2 f(\theta)\mathrm{d}\theta, \\
= \quad & \int_0^t \left[(\underline{y}-y)(\underline{y}+y) - 2(\underline{y}-y)(\theta-\beta)\right] f(\theta)\mathrm{d}\theta, \\
= \quad & 2(\underline{y}-y)T(t),
\end{aligned}
$$

which is strictly positive. Including y in the delegation set therefore strictly increases the principal's payoff, which implies $\min D^*(Y) = \min Y$. □

Lemma 13. *Let condition (7) be satisfied. The optimal delegation set is such that*

$$\max D^*(Y) = \arg\min_{y\in Y} |y - \hat{y}|.$$

Proof. Consider delegation set D and suppose $\max D < \max Y$. Let $\overline{y} = \max D$ and consider action $y > \overline{y}, y \in Y$. Let $t = \frac{y+\overline{y}}{2}$ denote the state at which the agent is indifferent between the two actions. The change in the principal's payoff when including action y is given by

$$\begin{aligned} & -\int_t^1 (y-\theta+\beta)^2 f(\theta)\mathrm{d}\theta + \int_t^1 (\overline{y}-\theta+\beta)^2 f(\theta)\mathrm{d}\theta, \\ = & -\int_t^1 \left[(y-\overline{y})(y+\overline{y}) - 2(y-\overline{y})(\theta-\beta)\right] f(\theta)\mathrm{d}\theta, \\ = & -2(y-\overline{y})S\left(t\right). \end{aligned}$$

This change is weakly positive if and only if $S(t) \leq 0$ and hence if and only if $t \leq \hat{y}$. Since t is the midpoint of $\overline{y}$ and y, this condition holds if and only if the distance between $\overline{y}$ and $\hat{y}$ is weakly greater than the distance between y and $\hat{y}$, i.e. $|\overline{y}-\hat{y}| \geq |y-\hat{y}|$. □

The previous results together conclude the proof. □

B.2 Proof of Proposition 4

Proof. Let $\hat{y}(\beta)$ be the optimal cap under full awareness when the bias is β. Hence, $\hat{y}(\beta)$ describes the solution of the problem

$$\max_y - \int_{y_{min}}^{y} \beta^2 dF(\theta) - \int_y^1 \left(y-(\theta-\beta)\right)^2 dF(\theta). \tag{23}$$

and is implicitly defined by the first-order condition

$$\int_{\hat{y}}^1 (\theta-\beta)\,\mathrm{d}F(\theta) - \hat{y}(1-F(\hat{y})) = 0.$$

The following second-order necessary condition must be satisfied at $\hat{y}(\beta)$:

$$\beta f(\hat{y}(\beta)) - (1-F(\hat{y}(\beta))) \leq 0. \tag{24}$$

By condition (7), the third derivative of expression (23) with respect to $\bar{y}$, given by $\beta f'(\bar{y})+f(\bar{y})$, is strictly positive. Hence, the first derivative is strictly convex. Strict convexity of the first derivative, together with the fact that it equals zero from below at $\bar{y} = 1$, implies that at $\hat{y}$ the first derivative crosses zero from above. Condition (28) must hence be satisfied with strict inequality. We can then use the implicit function theorem (condition (7) implies that the cumulate F is $\mathcal{C}^1$) to show that $\hat{y}(\beta)$ admits a derivative at each β, which equals:

$$\hat{y}'(\beta) = -\frac{1-F(\hat{y}(\beta)}{\beta f(\hat{y}(\beta)) - (1-F(\hat{y}(\beta)))} < 0 \tag{25}$$

where we used the necessary second order condition (28) with strict inequality. Continuous differentiability is guaranteed by the implicit function theorem and can be checked directly in the above expression

Turning to the case of partial awareness, let us first write the agent's payoff as a function of Δ and the parameter β:

$$U(\Delta;\beta) = -\int_{\hat{y}(\beta)-\Delta}^{\hat{y}(\beta)} (\hat{y}(\beta)-\Delta-\theta)^2 \mathrm{d}F(\theta) - \int_{\hat{y}(\beta)}^{1} (\hat{y}(\beta)+\Delta-\theta)^2 \mathrm{d}F(\theta). \tag{26}$$

The first and second derivative of this function with respect to Δ are

$$U'_{\Delta}(\Delta;\beta) = 2\int_{\hat{y}(\beta)-\Delta}^{\hat{y}(\beta)} [\hat{y}(\beta)-\Delta-\theta]\,\mathrm{d}F(\theta) - 2\int_{\hat{y}(\beta)}^{1} [\hat{y}(\beta)+\Delta-\theta]\,\mathrm{d}F(\theta) \tag{27}$$

$$U''_{\Delta\Delta}(\Delta;\beta) = -2[1-F(\hat{y}(\beta)-\Delta)] < 0. \tag{28}$$

Strict concavity implies that the agent's optimisation problem has a unique solution on $[0,\bar{\Delta}(Y)]$. The interior solution of this problem is characterised by the first-order condition

$$\int_{\hat{y}(\beta)-\Delta^*}^{\hat{y}(\beta)} [\hat{y}(\beta)-\Delta^*-\theta]\, f(\theta)\mathrm{d}\theta - \int_{\hat{y(\beta)}}^{1} [\hat{y}(\beta)+\Delta^*-\theta]\, f(\theta)\mathrm{d}\theta = 0. \tag{29}$$

Since F is $\mathcal{C}^1$, $\hat{y}(\beta)$ is $\mathcal{C}^1$ and $U''_{\Delta,\Delta} < 0$, the conditions for applying the implicit function theorem are satisfied. There is hence a function $\Delta^*(\beta)$ describing the unrestricted solution for the agent that solves the first order condition: $U'_{\Delta}(\Delta^*(\beta);\beta) = 0$, which becomes an identity when seen as a function of β, and:

$$\Delta^{*\prime}(\beta) = -\frac{U''_{\Delta,\beta}(\Delta^*(\beta);\beta)}{U''_{\Delta,\Delta}(\Delta^*(\beta);\beta)}.$$

To prove the statement of the proposition, we must then show $U''_{\Delta,\beta}(\Delta^*(\beta);\beta) > 0$. Differentiating the expression of the first order condition (10) with respect to β keeping Δ^* as fixed, after some rearrangement, delivers:

$$U''_{\Delta,\beta}(\Delta^*(\beta);\beta) = -\hat{y}'(\beta)\left[1 + F\left(\hat{y}(\beta)-\Delta^*(\beta)\right) - 2F(\hat{y}(\beta))\right].$$

Since $\hat{y}'(\beta) < 0$ (see (25)), we would be done if $1 + F\left(\hat{y}(\beta)-\Delta^*(\beta)\right) - 2F(\hat{y}(\beta)) > 0$. Note that this inequality can be equivalently written as:

$$2(1-F(\hat{y}(\beta))) > 1 - F\left(\hat{y}(\beta)-\Delta^*(\beta)\right).$$

Using $\hat{y}(\beta) = \mathbb{E}[\theta-\beta|\theta \geq \hat{y}(\beta)]$, the first order condition (29) can be written as:

$$[1-F(\hat{y}(\beta)-\Delta^*(\beta))]\left[\mathbb{E}[\theta|\theta \geq \hat{y}(\beta)-\Delta^*(\beta)] - (\hat{y}(\beta)-\Delta^*(\beta))\right] = 2[1-F(\hat{y}(\beta))]\beta. \tag{30}$$

Since $\hat{y}(\beta)-\Delta^*(\beta)$ is strictly smaller than $\hat{y}(\beta)$, the following condition holds:

$$\mathbb{E}\left[\theta-\beta|\theta \geq \hat{y}(\beta)-\Delta^*(\beta)\right] - \left(\hat{y}(\beta)-\Delta^*(\beta)\right) > 0.$$

Equivalently we can write

$$\mathbb{E}[\theta|\theta \geq \hat{y}(\beta) - \Delta^*(\beta)] - (\hat{y}(\beta) - \Delta^*(\beta)) > \beta.$$

Given this inequality, (30) requires $(1 - F(\hat{y}(\beta) - \Delta^*)\beta))) < 2(1 - F(\hat{y}(\beta)))$, as desired. $\square$

B.3 Proof of Proposition 8

Proof. We want to show that revealing an awareness set of the form $[y^A(0), \hat{y} - \Delta] \cup \{\hat{y} + \Delta\}$ is optimal. Towards a contradiction, suppose this is not the case and let the optimal awareness set be denoted by Y. Define $\tilde{\Delta}$ to be the largest value of Δ such that $Y \subseteq (-\infty, \hat{y} - \Delta] \cup [\hat{y} + \Delta, +\infty)$ and $\tilde{Y} := [y^A(0), \hat{y} - \tilde{\Delta}] \cup \{\hat{y} + \tilde{\Delta}\}$. Suppose the principal's realised awareness set is Y^P. According to Proposition 3, the induced delegation sets from revealing, respectively, Y and $\tilde{Y}$ are

$$\begin{aligned} D^*(Y \cup Y^P) &= \{y \in Y \cup Y^P : y \leq \arg \min_{y \in Y \cup Y^P} |y - \hat{y}|\}, \\ D^*(\tilde{Y} \cup Y^P) &= \{y \in \tilde{Y} \cup Y^P : y \leq \arg \min_{y \in \tilde{Y} \cup Y^P} |y - \hat{y}|\}. \end{aligned}$$

In order for Y to yield a strictly higher payoff for the agent than $\tilde{Y}$, there must exist some awareness set Y^P and some action y such that $y \in D^*(Y \cup Y^P)$ and $y \notin D^*(\tilde{Y} \cup Y^P)$. By its definition, $D^*(\tilde{Y} \cup Y^P)$ includes all actions in $\tilde{Y}$ weakly smaller than $\hat{y}$. Given $Y \subseteq \tilde{Y}$, it follows that $y > \hat{y}$. Furthermore, the optimal delegation set includes at most one action strictly greater than $\hat{y}$. By definition of $\tilde{\Delta}$, the set Y includes an action whose distance to $\hat{y}$ is $\tilde{\Delta}$. This implies that the largest action in $D^*(Y \cup Y^P)$ is weakly smaller than $\hat{y} + \tilde{\Delta}$. Hence, we have $y \leq \hat{y} + \tilde{\Delta}$. Also, since y belongs to $D^*(Y \cup Y^P)$, it follows that there is no action in Y^P strictly closer to $\hat{y}$ than y. However, the property $|y - \hat{y}| \leq \tilde{\Delta}$, together with the fact that there is no action in Y^P that is closer to $\hat{y}$ than y, implies that y must also belong to $D^*(\tilde{Y} \cup Y^P)$. A contradiction.

To prove the second part of the statement, let $EU(\Delta)$ denote the agent's expected payoff associated to the disclosure of a set of actions parametrised by Δ and let μ be the probability which the agent assigns to the event that the principal's awareness set does not include $\hat{y}$. Then, given our assumption that awareness sets are closed, there exists some $\varepsilon > 0$ such that no action in $(\hat{y} - \varepsilon, \hat{y} + \varepsilon)$ belongs to any of the realisations of the principal's awareness sets in the support that do not contain $\hat{y}$. For $\Delta \in [0, \varepsilon]$, the agent's expected payoff conditional on facing a principal who is not aware of $\hat{y}$ is then described by the function $U(\Delta)$, as defined in (26), with $U'(0) = -2 \int_{\hat{y}}^1 (\hat{y} - \theta) \, \mathrm{d}F(\theta) > 0$ (see (27))

With the complementary probability $1 - \mu$, the agent faces a principal who is aware of $\hat{y}$. In this event, the principal never permits an action greater than $\hat{y}$. A lower bound for the agent's payoff conditional on the principal being aware of $\hat{y}$ as a function of Δ is given by the payoff that obtains when the principal is unaware of all actions in $(\hat{y} - \Delta, \hat{y})$: any action in the principal's awareness set belonging to $(\hat{y}-\Delta, \hat{y})$ will be included in the delegation set and thus increases flexibility for the agent. The lower bound utility is:

$$\underline{U}(\Delta) = -\int_{\hat{y}-\Delta}^{\hat{y}-\Delta/2} (\hat{y}-\Delta-\theta)^2 f(\theta)\mathrm{d}\theta - \int_{\hat{y}-\Delta/2}^{1} (\hat{y}-\theta)^2 \mathrm{d}F(\theta),$$

with

$$\underline{U}'(\Delta) = 2\int_{\hat{y}-\Delta}^{\hat{y}-\Delta/2} (\hat{y}-\Delta-\theta)\mathrm{d}F(\theta)$$

The agent's unconditional expected payoff $EU(\Delta)$ must then be weakly greater than $\mu U(\Delta) + (1-\mu)\underline{U}(\Delta)$. Since $\underline{U}'(0) = 0$ and $U'(0) > 0$, the first derivative of $\mu U(\Delta) + (1-\mu)\underline{U}(\Delta)$ evaluated at $\Delta = 0$ is strictly positive. Given that $\mu U(\Delta) + (1-\mu)\underline{U}(\Delta)$ is equal to $EU(\Delta)$ at $\Delta = 0$ and that $\mu U(\Delta) + (1-\mu)\underline{U}(\Delta)$ constitutes a lower bound for $EU(\Delta)$ elsewhere, it follows that $EU(\Delta)$ is strictly increasing on a right neighbourhood of $\Delta = 0$.

□

C Generalised extensive-form games with unawareness

Heifetz et al. (2013, from now on HMS) define generalised extensive-form games that allow for evolving unawareness.

To introduce the generalised extensive-form game Γ, let N be a set of decision nodes, C be a set of chance nodes, and Z be a set of terminal nodes. The nodes $\bar{N} = N \cup C \cup Z$ constitute a tree. HMS capture limited awareness via the notion of subtrees, defined as subsets of nodes of $\bar{N}$. Letting $\mathbf{T}$ be a family of subtrees of $\bar{N}$, for $T, T' \in \mathbf{T}$ the relation $T' \preccurlyeq T$ signifies that the nodes of T' constitute a subset of the nodes of T. One element of $\mathbf{T}$ represents the modeler's view of the paths of play that are objectively feasible. The other elements of $\mathbf{T}$ represent feasible paths of play as subjectively viewed by some player, or as the frame of mind attributed to a player by other players or by the same player at a later stage of the game.

HMS propose a number of natural properties for generalised extensive-form games. These include basic extensions of standard requirements of extensive form games but also new properties that are specific to unawareness. All of these features are satisfied in our generalised game, which we describe in more detail now.

C.1 'Pure' Delegation Generalised Extensive-form Game

Consider first the model of the initial part of the paper where renegotiation is not possible. There are are three players: the principal, the agent, and nature/chance. Each game tree in $\mathbf{T}$ is very similar and all have virtually the same structure. An example of trees in $\mathbf{T}$ for the 'pure' delegation model is reported in Figure 3. At the root, the agent moves with $Y \supseteq Y^p$, then the principal proposes $D \subseteq Y$, then nature picks θ which is revealed to the agent,x and finally the agent picks $y \in D$ knowing θ. The agent's view is the richest one and coincides with the modeller's view and with the principal's view in case of full awareness. Let the associated game tree be denoted by T_{Y^A}. The other elements of $\mathbf{T}$ represent feasible paths of play as subjectively viewed by the principal. They also coincide with the way the agent see the principal's view. We denote by T_Y the subtree associated to principal's awareness level Y. The tree T_Y can be depicted starting from T_{Y^A} and deleting all moves Y' such that $Y \cap Y' = \emptyset$ or $Y \subset Y'$ (or both) and all nodes following those moves.

At the outset of the game, the principal's awareness is Y^P and her subjective view of the game is described by a subtree T_{Y^P}. Within the confined view of T_{Y^P} the principal can only envisage the agent announcing Y^P. If the agent reveals additional projects to the principal—i.e., he announces $Y \supset Y^P$—the principal updates her awareness and therefore her subjective view of the game. Given updated awareness Y, the principal's subjective game tree T_Y includes additional nodes. Once the principal becomes aware of the additional nodes, she can also contemplate less expressive game trees $T_{Y'}, Y^P \subseteq Y' \subseteq Y$. That is, the principal can envisage how the game would have unfolded had the agent revealed fewer actions. However, the principal cannot contemplate the paths of play that would have been feasible if the agent would have revealed more. After the initial stage, awareness no longer changes, which means that, given the constraints unawareness imposes on the players' succeeding moves, agent and principal play a standard game. Finally, it might be useful to note that, within the HMS structure, while the agent knows that the 'true' game tree is T_{Y^A}, the tree T_Y represents how the agent's thinks the principal sees the game when her awareness level is Y. Furthermore, the tree T_Y represents both how the principal actually sees the game when her awareness level is Y (so the agent is right) and what the principal thinks the agent's view of the game is (so the principal is wrong). The tree T_Y also represents all the higher order beliefs about each other views of both the agent and the principal.

Information Sets and Strategies. The 'forest' of trees $\mathbf{T}$ is hence constituted by $\mathbf{T} = \Big(\{T_Y\}_{Y \in \mathcal{Y}(Y^P)}\Big)$, where T_{Y^A} is the objective tree and, recall, $\mathcal{Y}(Y^P)$ is the set of all closed sets in $\mathbb{R}$ that are subset of Y^P.

A complete description of the strategic interactions among the players however requires considerations across the trees. The delegation game is of perfect information so for each player each information set contains a sigle node. The same node however, might appear in several trees. When a node is a decision node we treat the corresponding node in a different tree as a different node—we thereby generate copies of it—while keeping the obvious order of the nodes among all copies. Some nodes, however, do not constitute information sets of any player. So there is a redundancy in this description.

Information sets represent both information and awareness. In each tree T_Y, *the principal has exactly one information set*, where she decides the delegation set given his awareness. However, as explained above, given her awareness Y, the principal can now contemplate what she could have done under lower awareness. These views are described by nodes in T_Y that lead to information sets in less expressive trees: $T_{Y'}$ for $Y^P \subseteq Y' \subseteq Y$. The nodes representing lower awareness than Y in T_Y are reported in T_Y but they do not constitute principal's information sets as they do not represent the 'state of mind' the principal would have at that node. The *agent's information sets* are the root in each T_Y and the nodes (Y', D, θ) for $Y' \subseteq Y$, $D \subseteq Y$ an $\theta \in [0, 1]$.

A T_Y*-partial game* in our framework is constituted by the partially ordered set of trees including T_Y and all $T_{Y'}$ with $Y' \subseteq Y$, with all the information sets as specified above. The T_{Y^A} together with the linked less expressive trees and the information sets of the two players specifying subsets of the nodes in $\mathbf{T}$ constitutes the *generalised extensive-form game.*

The *agent's pure strategy* restricted to the T_Y-partial game can be described by a collection of pairs $\sigma_{Y'}$ and $y_{Y'}$, one pair for each $Y^P \subseteq Y' \subseteq Y$, where $\sigma_{Y'}$ represents the move at the root of the tree $T_{Y'}$ with feasibility $Y^P \subseteq \sigma_{Y'} \subseteq Y'$ and the function $y_{Y'}(\cdot)$ maps one $y \in D$ for each $\sigma_{Y'} \subseteq Y', \theta \in [0, 1], D \subseteq \sigma_{Y'}$; that is, $y_{Y'}(\sigma_{Y'}, \theta, D) \in D$. The subscript in the functions

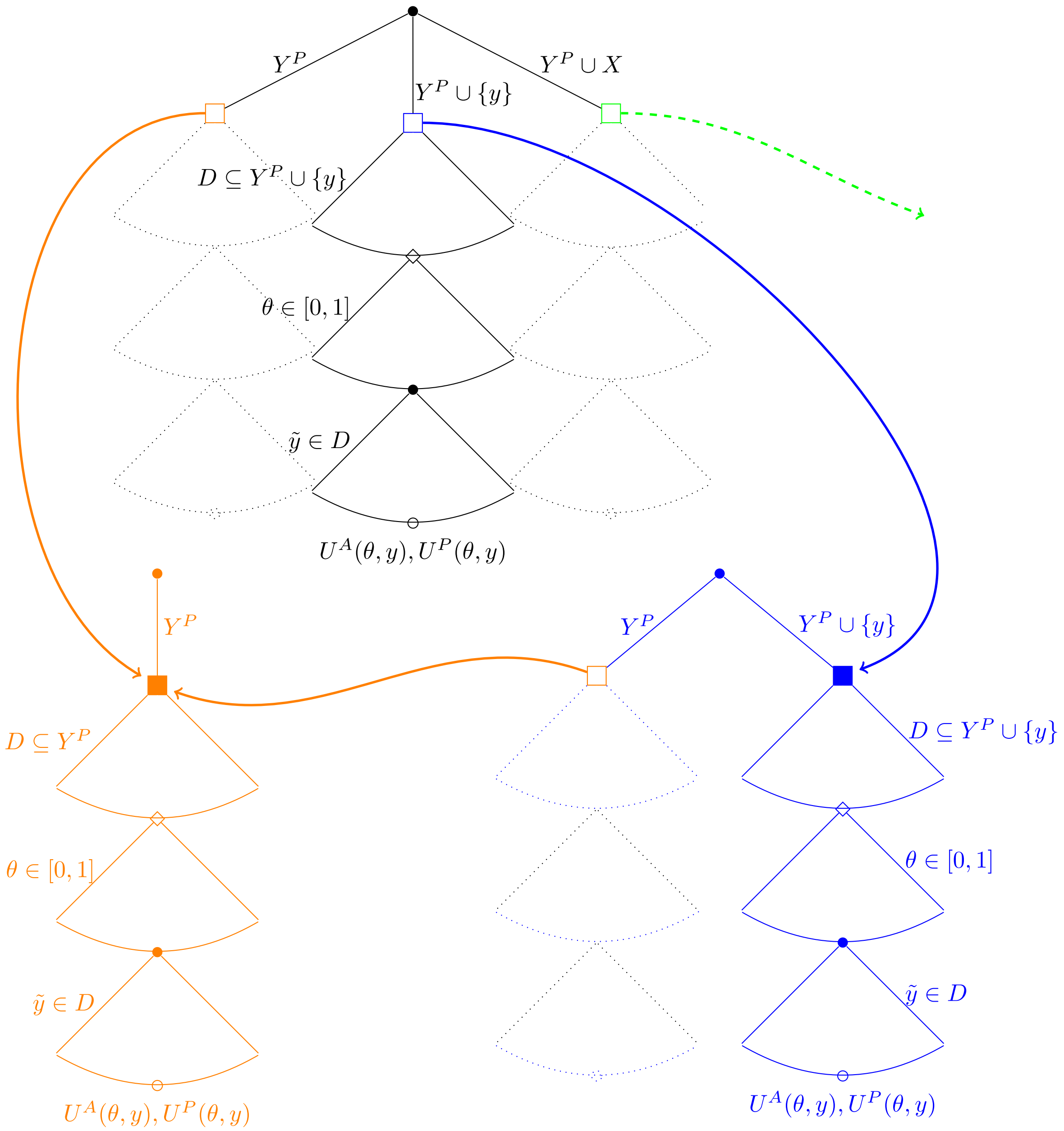

Figure 3: Objective and subjective game trees: circles are decision nodes of the agent, diamonds those of nature, and squares those of the principal. The tree on the top depicts the objective game tree. The tree on the bottom left illustrates the principal's subjective view of the game before the agent discloses additional actions. The tree on the bottom right shows the principal's subjective view of the game after becoming aware of action y. In the branches of the different trees we also indicate the principal's awareness. After updating the principal can contemplate how the game would have unfolded if she had not learned about y but she cannot contemplate branches of the objective game tree following disclosures of actions different than y. Solid squares (resp. circles) denotes the principal's (resp. agent's) information sets. The principal has exactly one information set in each of the bottom trees.

indicates the tree where the nodes constituting agent's information sets are located.

Recall that to each node $Y' \subset Y$ in T_Y we associated an information set (and hence a decision node) for the principal in tree $T_{Y'}$ (i.e., to a tree outside T_Y). The structure of $\mathbf{T}$ is relatively simple in that the less expressive trees are linked to more expressive ones in a simple way. This simplicity allows us to describe *a pure strategy for the principal* as a function $D(\cdot)$ that associate to each $Y \in \mathcal{Y}(Y^P)$ a set in $\mathcal{D}(\mathcal{Y})$.[28] Actions taken by the principal at a given informations set also generate the same moves at the corresponding nodes in trees with higher levels of awareness. Each function $D(\cdot)$ embeds all information for a complete description of the principal's strategy in the generalised game. In particular, given a function $D(\cdot)$, we can describe the principal's strategy restricted to a T_Y-partial game by considering the behaviour of D restricted to the domain $Y' \subseteq Y$.

C.2 Rationalizability and Prudent Rationalizability

In this section we show that *the equilibrium described in the main text is the unique* Δ-*rationalizable outcome of the generalised game we described above whenever we restrict to pure strategies and assume the two tie-breaking rules reported in the main text to be commonly believed.*

System of Beliefs A system of beliefs associates to each information set h of a player, his/her beliefs over the opponent's strategies restricted to the $T-$partial game that includes the information set, assigning probability one to the opponent's strategies that lead to the information set h. The system is 'prudent' whenever it gives otherwise full support to all 'conceivable' strategies of the opponent (see Heifetz et al. (2020) for details).

First, we describe the *principal's set of beliefs.* When confronted with awareness Y, the principal must attach probability one to all strategies of the agent that include the first move $\sigma_{Y'} = Y$ (so only to the node $\sigma_Y = Y$ in the tree T_Y) and then must form beliefs about the continuation of the agent's strategy $y_{Y'}$ in the last stage for all $Y' \subseteq Y$. However note, that the entry Y in the functions describing the component of the agent's strategy at the last stage is only in trees T_Y or more expressive trees. But such more expressive trees are not part of the T_Y-partial game. This implies that the principal must assign probability zero to all such functions. Moreover, since the agent's move at the root is taken before knowing the realisation of θ, all of the equilibrium considerations are useless for the principal in terms of inference based on forward induction considerations. As a consequence, what matters for the principal's beliefs at the (sole) information set in tree T_Y is the agent's strategy y_Y in that tree alone. This is what we will keep track of. So, our 'prudent rationalizability' sets omit some (redundant and irrelevant) elements compared to the defintion by Meier and Schipper (2014).

[28]Since the principal has one information set in each tree, we could have equivalently written her strategy as an action D_Y in the sole information sets for the tree T_Y.

In our problem, the only *agent's beliefs* that need to be specified are those at the root about the principal's rationalizable strategies $D(\cdot)$. Those refer to actions taken over different trees. However, as argued above, they can be described by a single function.

To define prudent rationalizability, consider hence the following sequence of sets:

$$\begin{aligned} R_0^A(Y) &= \left\{(\sigma_Y, y_Y): \ \sigma_Y \in \mathcal{Y}, \sigma_Y \subseteq Y, \text{ for } Y' \subseteq Y, D \subseteq Y', \ y_Y : (Y', \theta, D) \mapsto D\right\}, \\ R_0^P &= \left\{D: \ \forall Y \in \mathcal{Y}, \ D(Y) \in \bar{\mathcal{D}}(Y)\right\}, \end{aligned}$$

and for $k > 0$:

$$R_k^A(Y) = \left\{(\sigma_Y, y_Y) \in R_{k-1}^A(Y) : y_Y(Y', \theta, D) \in BR^A(\theta, D) \ \exists \mu^A \in \bar{\Delta}(R_{k-1}^P); \sigma_Y \in \arg\max_{\hat{Y} \supseteq Y^p} \int_{R_{k-1}^P} V^A(\hat{Y}, D(\hat{Y}), y_Y) d\mu^A\right\},$$

$$R_k^P = \left\{D \in R_{k-1}^P : \forall Y \in \mathcal{Y} \ \exists \mu^P \in \bar{\Delta}(R_{k-1}^A(Y)); \ D(Y) \in \arg\max_{\hat{D} \subseteq Y} \int_{R_{k-1}^A(Y)} V^P(Y, \hat{D}, y_Y) d\mu^P\right\}.$$

In the above:

1. The entry Y in $R^A(\cdot)$ indicates strategies defined on information sets located in the tree T_Y. Recall that those are the only ones we will keep track of in the T_Y-partial game.

2. $\bar{\Delta}(R_{k-1}^A(Y))$ is the set of distributions over $R_{k-1}^A(Y)$ that gives probability one to $\sigma_Y = Y$ and has 'full support' over all functions y_Y in $R_{k-1}^A(Y)$ that satisfy the tie-breaking rule;

3. $\bar{\Delta}(R_{k-1}^P)$ is the set of distributions over R_{k-1}^P with 'full support' over the delegation sets in it that satisfy the tie-breaking rule;

4. The set $BR^A(\theta, D)$ indicates the best response of the agent as defined in (1).

5. For $i = A, P$, $V^i(Y, D, y_Y) := \int_0^1 U^i(\theta, y_Y(Y, \theta, D)) dF(\theta)$.

The sets of prudent-rationalizable strategies for player $i = A, P$ are defined as:

$$R_\infty^A(Y) := \cap_{k=0}^\infty R_k^A(Y) \qquad \text{and} \qquad R_\infty^P := \cap_{k=0}^\infty R_k^P.$$

The technical complications involved in the definition of 'full support' are not analysed in detail because, as we will see next, the sets quickly shrink to singletons.

Consider indeed the sets $R_1^A(\cdot)$ and note that R_0^P allows for all $D \subseteq Y$ in each information set $\sigma = Y$, so all agent's information sets are reached. Furthermore, it is immediate to see that the objective function and the feasibility sets defining y_Y are identical across all sets. With an appropriate restriction of the domain (to be expressible within the tree T_Y) each of these rules can hence be described concisely by a function describing the agent's best strategy in all trees. The function y_Y would then coincide with this extended function when its domain is restricted to $Y' \subseteq Y$ and $D \subseteq Y'$. Moreover, the commonly believed tie-breaking rule implies that such function is unique. Recalling the notation used in Appendix A.3, let's denote this function by y^*.

Next, while R_1^P is still a rich set, we now argue that R_2^P is a singleton. The argument in the previous paragraph implies that for each Y, the belief μ^P must be concentrated on the singleton $(\sigma_Y, y_Y) = (Y, y_Y^*)$ (where we denote by y_Y^* the reduced domain version of y^*). This implies that, for each Y, when deciding D, the principal only considers the objective function $V^P(Y, D, y_Y^*)$ (with full probability weight on this combination alone). This, together with the tie-breaking rule implies that the principal has only one solution for each Y. This is what we denote in the main text as $D^*(Y)$. The solutions across the principal's information sets defines the function $D^*(\cdot)$

Finally, consider the sets $R_3^A(\cdot)$. We already argued that the last stage component can be described by the function y^*. Moreover, we argued in the previous paragraph that R_2^P is a singleton. The objective of the agent is hence $V^A(Y, D^*(Y), y_Y^*)$ with full weight on this combination alone.

Recalling that for all Y, D, θ we have $y_Y^*(Y, D, \theta) = y^*(Y, D, \theta)$, our findings can be summarised as follows:

Proposition 14 (Rationalizable Outcome)**.** *Let Y^P given. The 'relevant' set of (prudently) rationalizable strategies can be compactly defined as the sets (R^A, R^P) satisfying:*

$$\begin{aligned} R^P &= \left\{ D^* : \forall Y \in \mathcal{Y},\ D^*(Y) \in \arg\max_{\hat{D} \subseteq Y} V^P(Y, \hat{D}) \text{ together with the tie-breaking rule} \right\}, \\ R^A &= \left\{ (Y, y^*) : \ y^*(Y, \theta, D) \in BR(\theta, D) \text{ with tie-breaking rule},\ \ Y \in \arg\max_{\hat{Y} \supseteq Y^p} V^A(\hat{Y}, D^*(\hat{Y})) \right\}, \end{aligned}$$

where for all $Y \in \mathcal{Y}$ and $D \in \bar{\mathcal{D}}(Y)$:

$$V^A(Y, D) := \int_0^1 U^A(\theta, y^*(Y, \theta, D)) dF(\theta); \tag{31}$$

$$V^P(Y, D) := \int_0^1 U^P(\theta, y^*(Y, \theta, D)) dF(\theta). \tag{32}$$

The proof is now complete, it is indeed immediate to see that the singletons described in R^P and R^A coincide with the Perfect Bayesian Nash equilibrium strategies described in the main text.

C.3 The Generalised Game with Renegotiation

In this subsection, we describe in detail the generalised extensive-form game representing our model with the possibility of renegotiation. An example of the forest of trees is reported in Figure 4.

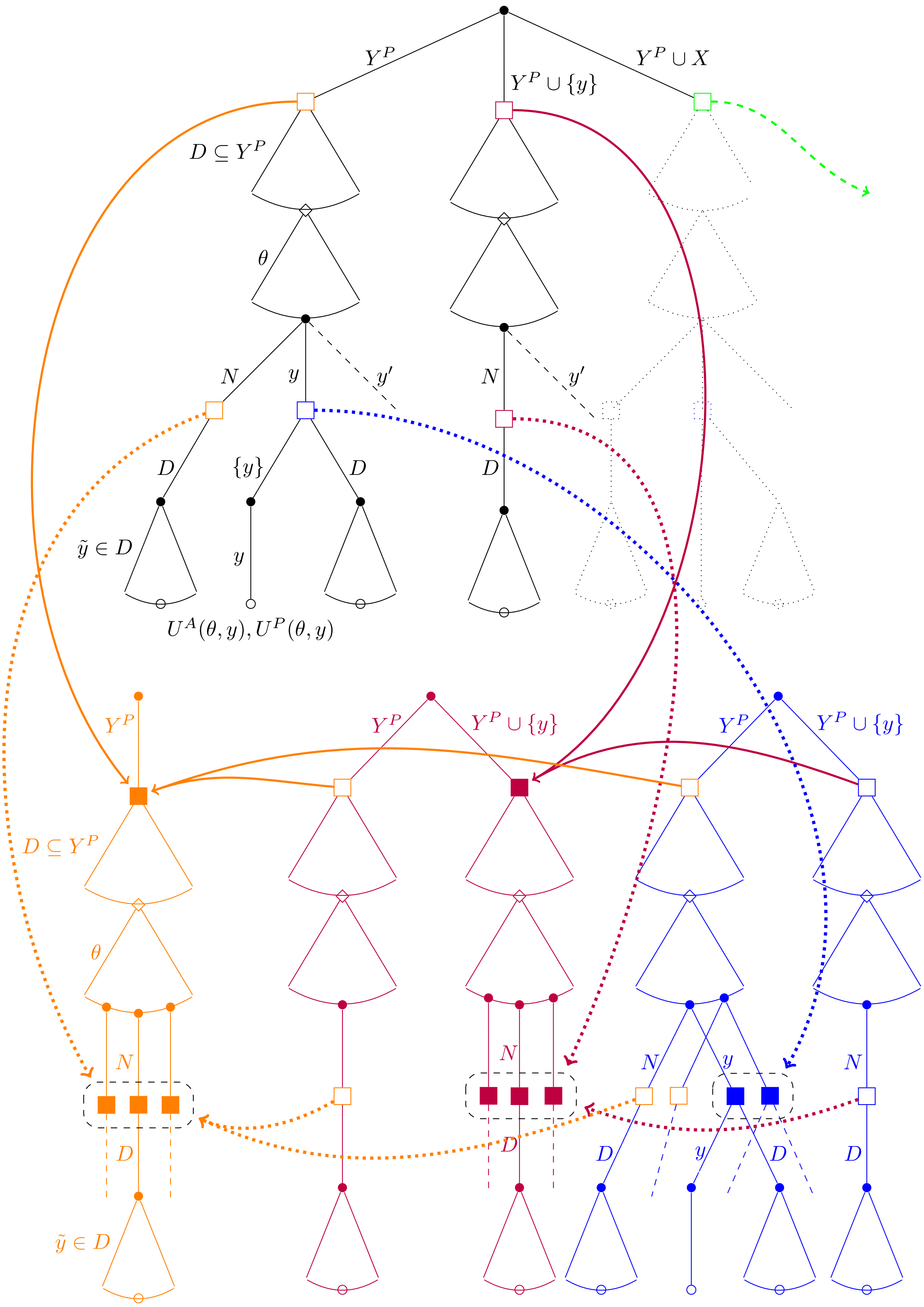

Figure 4: Objective and subjective trees for the game with renegotiation.

There are two instances in the generalised game where the principal may change her view. The first one is after the initial move by the agent, as in the 'pure' delegation game we described in Section C.1. A second instance where the principal may change her awareness is at the renegotiation stage, whenever the agent proposes an action outside the awareness of the principal. The generalised game is hence constituted by more trees compared to the 'pure' delegation case.

Recall that we assume that the principal is initially unaware of the possibility of renegotiation. This assumption, first of all, implies that the generalised game includes trees very similar to the 'pure' delegation case. They represent situations where the principal only becomes aware of new actions right after the agent's move at the root and there is no proposal at the renegotiation stage. To be consistent with the previous notation, we denote all such threes as T_Y. However, to be consistent with a richer continuation of the tree in the renegotiation stage, we append to them a singleton move by the agent and a singleton move by the principal, the agent's move can be seen as an 'articificial' no-proposal which is followed by the principal's move assigning again the original D. These moves lead to a final stage that is identical to that in the 'pure' delegation case, where the agent can pick any $y \in D$. The tree T_Y differs from the tree that would represent the view of the principal who has been made aware of Y at the initial stage and does not receive any proposal outside her awareness set at the renegotiation stage, but she is aware of the possibility of renegotiation. The trees representing the latter situation/view are denoted by $T_{Y,\emptyset}$, where the second entry in the subscript indicates that there is no proposal *outside the awareness set at the renegotiation stage*, so the principal's awareness over the set of actions is still Y, although, in this case, she is aware of the possibility or renegotiation. The notation also permits to compute the final awareness of the principal in the tree: in this case, $Y \cup \emptyset = Y$.

Recall further that the agent may only propose singletons at the renegotiation stage. Consider awareness level Y after the agent's move at the root, and a proposal $y \notin Y$ at the renegotiation stage. This leads to a principal's information set that belongs to the game tree denoted as $T_{Y,y}$. In this tree, the principal contemplates the fact that at the beginning of the game the agent could have revealed $\sigma = Y \cup \{y\}$ (followed by N at the renegotiation stage) or any subset of $Y \cup \{y\}$ which is a superset of Y^P. Similarly, the principal realises that the agent could have revealed any other $y' \in Y$ at the renegotiation stage had he started at the root with $\sigma' = Y' \subset Y$, $\{y\} \cap Y' = \emptyset$, and so on. Our notation implies that the tree representing both the modeller's and the agent's view is denoted by $T_{Y^A,\emptyset}$. This tree will, of course, be identical to all other trees where the principal has full awareness and she is also aware of the renegotiation stage. As anticipated, with the exclusion of the trees T_Y, where the principal remains unaware of the possibility of a renegotiation stage, all other trees are identical as long as the principal's final awareness is the same.[29] Despite these similarities, in order to describe the principal's information sets, all such trees must be treated separately. This is because within a generalised

[29]For example, the tree $T_{Y \cup \{y_1\},y_2}$ is identical to the tree $T_{Y \cup \{y_2\},y_2}$, because in both games the principal is aware of the possibility of renegotiation and has the same overall awareness set.

game, information sets model both information and awareness.

Information sets and Strategies The forest of trees **T** is constituted by trees of the form (in increasing order of awareness): T_Y, $T_{Y,\emptyset}$, and $T_{Y,y'}$, for $Y \in \mathcal{Y}(Y^P)$ and $y' \in Y^A \backslash Y$.

In each tree of the form T_Y, the principal has a single information set after a specific move at the root by the agent: this is the move $\sigma_T = Y$. At these information sets the principal chooses the delegation set $D \subseteq Y$. In T_Y the principal has also one information set after the agent's 'artificial' proposal N at the renegotiation stage but the principal has a singleton move at this information set, so it is less relevant.

In each tree $T = T_{Y,y}$, the principal has information sets at the renegotiation stage alone. To describe the principal's information sets at the renegotiation stage, say, within the tree $T_{Y,y}$, recall that the agent's proposal y at the renegotiation stage occurs after he observes θ, while the principal cannot observe it. Moreover, the principal might have taken a different delegation set after the agent's first move. Given $D \subseteq Y$, a principal's information set in the tree $T_{Y,y}$ is constituted by all nodes of the form (Y, D, θ, y) for $\theta \in [0, 1]$. In other terms, in each one of the trees $T_{Y,y}$, the principal has several information sets, one for each $D \subseteq Y$, and each one of such information sets includes all nodes with the same agent's moves in the renegotiation stage (y in this case) and all $\theta \in [0, 1]$. The *agent has information sets* at each decision node across the different trees, and they are all singletons.

A principal's strategy in the generalised game is described by a move in each information set. Given the specific structure of the principal's information sets across the trees constituting the generalised game, a principal's strategy can be described by a delegation move D_T and a reply function ρ_T for each tree T of the form T_Y. The delegation move is as in the 'pure delegation' case: it assigns a delegation set D given the awareness of the principal at this stage of the tree. As indicated, the function ρ_T for the three T_Y is reported for completeness but it is a singleton choice.

In trees of the type $T_{Y,y}$, there is no information set for D_T while ρ_T takes the particular history (Y, D, y) in the tree and maps into $\{a, r\}$, where a stays for 'accept' and r for 'reject' the proposal.[30] Note that within the tree, only the entry D can vary. This is because each alternative move of the agent $Y' \neq Y$ or $y' \neq y$ would lead to a principal's information set in a different tree.[31] In particular, although the agent is aware of the possibility of renegotiation, the information set corresponding to history (Y, D, N) in the tree $T_{Y,y}$ lies in the tree T_Y.

A complete description of the agent's set of strategies requires the indexation by the tree and the history in each tree. Fix a tree T. Then σ_T represents the agent's move at the root of the tree, x_T the proposal function and y_T the agent's last stage choice. The function x_T takes tuples (σ_T, D, θ) into a proposal $x \in Y^A$, while the function y_T takes a tuple $(\sigma_T, D, \theta, x_T, D')$ to D', where $D' \in \{D, x_T\}$. All such functions must be consistent with being in tree T, which indicates

[30]Recall, when $\rho_T = a$ then the delegation set the agent faces is $D' = y$, and if $\rho_T = r$ then $D' = D$.

[31]If $Y' \subset Y$ and $y' \in Y$ these nodes are present in the tree as hypothetical decision nodes for the principal, but they do not constitute principal's information sets.

the view of the principal. So, for example, if $T = T_{Y,y}$ then $Y^P \subseteq \sigma_T \subseteq Y$, $x_T \in Y \cup \{y\}$, and of course $D \subseteq \sigma_T$ and $y_T \in D'$.

We now describe what a *T-partial game associated to a given information set* is. There are two main types of T-partial games. First of all, if the principal remains unaware of the possibility or renegotiation, her information sets lie into $T-$partial games constituted by T_Y and all less expressible trees (i.e., $T_{Y'}$ for all $Y' \subset Y$). Again, this is in full analogy to the 'pure' delegation case. We then have T-partial games that are associated (i.e., include) information sets in which the agent is aware of the possibility of renegotiation. If the information set is at node (Y, D, y) in the tree $T_{Y,y}$, the T-partial game associated to this information set is constituted by the following trees: (i) $T_{Y'}$ for all $Y' \subseteq Y \cup \{y\}$; and (ii) $T_{Y',y'}$ where $Y' \cup \{y'\} \subseteq Y \cup \{y\}$, including the tree $T_{Y \cup \{y\}, \emptyset}$.

In order to define a Bayesian equilibrium, we need to define beliefs over nodes at information sets that include more then one node. As described above, the only instance where this happens is at the principal's information sets at the renegotiation stage after the agent's proposal. Given the commonly known cumulate F, and the focus on pure strategies, the principal's beliefs at such information sets are uniquely defined by the support. The support's belief function will be denoted as $\Theta_T(\cdot)$ and, for each tree T it maps triplets (Y, D, x) into subsets of $[0, 1]$. As mentioned above, such beliefs are irrelevant for the principal's choice in all trees T_Y. Given that both the agent and the principal have a singleton choice, we set w.l.o.g. $\Theta_T(\cdot) = [0, 1]$ for such trees (both on and off the equilibrium).

The literature provides several definitions of (Perfect) Bayesian Equilibrium and Sequential Equilibrium for generalised extensive-from games with unawareness of actions. Some of them are defined over slightly different frameworks. This is for example the case of Halpern and Rêgo (2014) and Feinberg (2020).[32] Our signalling game of the renegotiation stage fits into the class of dynamic games studied by Ozbay (2007) in that there is a fully aware agent who moves first.

Definition 2. *A PBE equilibrium for the generalised game is constituted by an 'assessment', that is, a pair of strategies for the agent and the principal,*

$$\{(\sigma_T^*, x_T^*, y_T^*)\}_{T \in \mathbf{T}} \text{ and } (\{D_{T_Y}^*\}_{Y^P \subseteq Y \subseteq Y^A}, \{\rho_T^*\}_{T \in \mathbf{T}}),$$

together with a beliefs' support system $\{\Theta_T^*\}_{T \in \mathbf{T}}$ for the principal at the renegotiation stage, that satisfy the following conditions. (i) At each information set, each player's choice maximises his/her expected utility, given the equilibrium strategies, conditional on his/her awareness, and the associated beliefs' support system. (ii) The strategies and beliefs' support system constitute a PBE in each subtree within any $T-$partial game. (iii) The beliefs' support system is consistent with the principal's awareness and equilibrium strategies whenever possible, and the conditional probabilities over each element in Θ_T^* are computed using the cumulate F.

[32]For a related equilibrium definition for games with unawareness in normal-form see Meier and Schipper (2014).

The meaning of the qualification 'conditional on his/her awareness' in condition (i) is that the choices and beliefs at each information set h only consider payoffs and strategies in the T-partial game that include the information set h. Condition (ii) is imposed for consistency with the definition given in Appendix A.3 which requires PBE at each (Y, D).

The next proposition relates the equilibrium set according to the previous definition to the equilibrium set according to the definition provided in Appendix A.3.

Proposition 15. *(i) Let (σ^*, x^*, y^*), (D^*, ρ^*) and Θ^* be an equilibrium, where $(\sigma^*, D^*(\cdot))$ is the solution of the contracting phase given y^* and (x^*, ρ^*, y^*) together with Θ^* constitute PBE equilibria according to Definition 1 in Appendix A.3. Then we can find a PBE equilibrium according to Definition 2 that is payoff equivalent and, in fact, coincides in all 'payoff-relevant' elements.*
(ii) The payoff-relevant components of the agent's best equilibrium outcome are the same across the two definitions.

Proof. (i) Let us start with the function y_T, $T \in \mathbf{T}$ describing the agent's strategy in the very last stage. It is clear that, given (θ, D'), the last stage problem of the agent is identical across all trees. We can hence describe the agent's strategy component at the last stage with functions y that map tuples (Y, D, θ, D') into D'. Given y, each functions y_T can be recovered by appropriately reducing the domain of the function y. The structure of the generalised game allows for different functions in each tree. As we have argued above, once we impose the tie-breaking rule, the agent's final mapping compatible with equilibrium will be a single function. As a consequence, the restriction we impose here is w.l.o.g. as long as we require subgame perfection. In fact, abusing notation, we let y^* be such function and its components that lead to $D' = D$ as it has been defined in the pure delegation model, while for $D' = x$ we obviously have $y^*(\cdot, x) = x$.

Recall that the delegation choice of the principal is a move for each Y, so it can be described—exactly as in the 'pure delegation' case—by a function mapping principal's awareness Y into a delegation set $D \subseteq Y$. Moreover, given the unawareness about the renegotiation stage at all such information sets, the principal's considerations regarding the future at these information sets are summarised by the map y^*, as in the 'pure' delegation case. In our framework, to take into account the unawareness of the renegotiation phase, the functions which the principal sees are those where the last entry is $D' = D$, which in turn is the consequence of the (artificial) 'no proposal' we assumed at the renegotiations stages in all trees T_Y. The tie-breaking rule implies that the equilibrium delegation function is unique and it coincides with the optimal one in the 'pure delegation' model, which we denoted by D^*.

So far we have shown that both equilibrium definitions use the (unique) functions D^* and y^* of the pure delegation model (appropriately adjusted for the extensive form).

We now move to the ρ function component of the principal's strategy. We want to show that using $\rho^*_T = \rho^*$, as defined in Appendix A.3, for all trees where the principal is aware of the renegotiation stage (with appropriate adjustments for the domain) and trivial choices for trees where the principal is unaware of renegotiation, constitutes an equilibrium for the generalised game.

For trees corresponding to the principal's views where she is aware of the renegotiation stage, the ρ_T component of the principal's strategy will be determined by the principal's beliefs at the information set (Y, D, x) of the tree $T_{Y,x}$. Note that all changes in Y and x that lead to changes in principal's awareness lead to information sets in different trees, so the only redundancy comes from the principal's choice D. Since the agent's initial move σ is done before knowing θ, the principal's inference at node (σ_T, D, x_T) only depends on the pair (D, x), her awareness Y, and the agent's strategy at the node. Note that the principal's beliefs and choices off the equilibrium path may matter for the agent's choice at the root. It will be useful to distinguish two type of off-the-equilibrium beliefs. First, whenever the node has the D-component off-the-equilibrium, beliefs and the principal's choices do not affect the agent's choice. On the other hand, for nodes of the form $(Y, D^*(Y), x)$, for each $Y \supseteq Y^P$, we have shown in Proposition 5 that in each PBE at the renegotiation stage where x is accepted by the principal, the beliefs' support is:[33]

$$\Theta(Y, D^*(Y), x) := \{\theta \in [0,1] : U^A(\theta, x) \geq U^A(\theta, y'), \forall y' \in D^*(Y)\}. \tag{33}$$

Note in particular, that this set does not depend on the agent's strategy and that the set of the agent's proposals that can be accepted by the principal in a PBE is again defined by Proposition 5.

We have hence shown that for trees where the principal is aware of the renegotiation stage we can set $\rho^*_T = \rho^*$, where ρ^* is the equilibrium function according to Definition 1 (mapping tuples (Y, D, x) to $D' \in \{D, x\}$); and the principal's belief at node (Y, D, x) can be the set $\Theta^*(Y, D, x)$ of feasible θ, satisfying (33) together with the marginal induced by the cumulate F obtained from Definition 1. If faced with the entry (Y, D, N), or in trees where the principal is unaware of the renegotiation stage, the function ρ_T has a singleton choice $\rho_T(\cdot) = D$ and we can set $\Theta^*_T(Y, D, N) = [0, 1]$ at these information sets. Similarly, we can set $x^*_T = x^*$ for all trees where the principal is aware of renegotiation and $x^*_T = N$ for all other trees.

It is immediate to see that the mentioned strategies together with the initial choice $\sigma^*_T = Y^*$ in all trees reached in equilibrium for some realisation of θ constitute a PBE according to Definition 2. The agent's choice at the root of trees which are not reached in equilibrium will also be specified in Defintion 2, while they are not defined in the more compact definition of Appendix A.3.

Point (ii) is a direct consequence of the fact that in each equilibrium according to Definition 2, the set of the agent's proposals that can be accepted by the principal is characterised by Proposition 5. □

[33]It is easy to see that the inference leading to the set Θ^* holds in this context as well.