The Limits of Limited Commitment

Jacopo Bizzotto † Toomas Hinnosaar ‡ Adrien Vigier §

February 20, 2025

Abstract

We study limited strategic leadership. A collection of subsets covering the leader’s action space determines her commitment opportunities. We characterize the outcomes resulting from all possible commitment structures of this kind. If the commitment structure is an interval partition, then the leader’s payoff is bounded by her Stackelberg and Cournot payoffs. Under general commitment structures the leader may obtain a payoff that is less than her lowest Cournot payoff. We apply our results to a textbook duopoly model and elicit the commitment structures leading to consumer- and producer-optimal outcomes.

JEL: C72, D43, D82

Keywords: commitment, Stackelberg, Cournot

*We thank the editor, Alexander Wolitzky, as well as Giacomo Calzolari, Matteo Escudé, Dino Gerardi, Francesco Giovannoni, Aditya Kuvalekar, Ignacio Monzón, Balázs Szentes and three anonymous referees for their insightful comments.

†OsloMet, jacopo.bizzotto@oslomet.no.

‡University of Nottingham and CEPR, toomas@hinnosaar.net.

§University of Nottingham, adrien.vigier@nottingham.ac.uk.
1 Introduction

In many strategic settings, a player’s early decisions constrain the actions she will later be able to take. The Stackelberg leadership model provides the best-known example of this kind: one firm (the leader) chooses a quantity first; another firm (the follower) observes the decision of the leader, and then chooses its own quantity. Numerous variants of this model are widely used, in industrial organization (Tirole, 1988), macroeconomics (Kydland and Prescott, 1977; Barro and Gordon, 1983), computer science, operations research, and marketing. Yet, by allowing the leader to select a single action early on, the Stackelberg model supposes a very strong form of commitment. While this “full-commitment” assumption is undoubtedly warranted in a number of settings, many real-world situations are such that early decisions restrict—but do not strictly fix—a player’s subsequent choice. The term “limited commitment” captures the leader’s inability to commit to a single action of her choice.

Limited commitment can manifest itself in many ways. For instance, an incumbent firm could influence potential entrants by investing in the development of a new product, building a plant in a new country, or announcing a list price while leaving the door open to discounts. We thus propose a model which allows for arbitrary forms of limited commitment. We characterize all outcomes resulting from limited commitment, compare them to the full and no-commitment benchmarks, and thereby evaluate the robustness of predictions based on these benchmarks.

The following example illustrates the power of limited commitment. There are two firms, an incumbent and an entrant, respectively producing quantities $q_1$ and $q_2$; production costs are zero, and the inverse demand function is $1 - q_1 - q_2$. The full-commitment (or Stackelberg) outcome is such that the incumbent produces 1/2 whereas the entrant produces 1/4; the resulting price is 1/4. By contrast, the no-commitment (or Cournot) outcome is such that each firm produces 1/3, leading to the higher price 1/3. As it turns out, limited commitment may benefit consumers by inducing a price lower than in both of these cases. For instance, suppose that the only commitment options of the incumbent are (i) produce at most 2/3, (ii) produce at least 2/3. In that case, the incumbent chooses the second option and ends up
producing $2/3$. The entrant produces $1/6$, and the resulting price is $P = 1/6$.

The previous example is a special case of the general framework we study. In our model, there are two periods and two players, a leader and a follower. Players’ action spaces are closed intervals and their payoffs are quasi-concave in one’s own actions. The setting is parametrized by a collection of subsets that cover the leader’s action space; we refer to this collection of subsets as the commitment structure. In the first period, the leader selects an element from the commitment structure. In the second period, leader and follower simultaneously choose one action each, the leader being restricted to pick an action from the subset that she selected in the first period. The full- and no-commitment benchmarks are special cases of our model: in the former, the commitment structure consists of singletons; in the latter, the commitment structure includes only one subset—the leader’s entire action space.

We characterize the set of outcomes that are part of an equilibrium under some commitment structure. We label those outcomes plausible. As illustrated in the example above, the set of plausible outcomes extends beyond the bounds defined by the full and no-commitment benchmarks. More strikingly, this remark applies to the players’ equilibrium payoffs too. In particular, although full commitment always benefits the leader, we show that the leader can be worse off under limited commitment than under no commitment. In fact, in a standard symmetric duopoly setting—where full commitment confers a well-known first-mover advantage—certain forms of limited commitment even generate a second-mover advantage.

We then examine the set of outcomes resulting from “simple” commitment structures, namely commitment structures which divide the leader’s action space into non-overlapping intervals. We call these outcomes simply plausible. We show that any outcome such that the follower best-responds and guaranteeing the leader at least her highest no-commitment payoff is simply plausible. Furthermore, any simply-plausible outcome must give the leader at least her lowest no-commitment payoff. Thus limited commitment always benefits the leader under simple commitment structures. However, the set of simply-plausible outcomes might exhibit “gaps,” in the sense that certain outcomes giving the leader a payoff between her full and

---

1 That is, players’ equilibrium actions need not be bounded by their full and no-commitment counterparts.
no-commitment payoffs may fail to be simply plausible.\textsuperscript{2}

The final section of our paper presents an application of our results to a textbook duopoly setting. We first characterize all plausible outcomes of that setting, and then describe the consumer surplus and social welfare-maximizing plausible outcomes. We show that, in terms of consumer surplus and total welfare, some forms of limited commitment are always strictly preferable to the traditional benchmarks of full and no commitment.

We contribute to a literature on commitment whose starting point is that economic agents often commit to \textit{subsets} of actions rather than single actions. This literature is divided in two branches. One branch posits that agents commit to subsets of actions because they are constrained to do so: even if they wanted to, agents would be unable to commit to specific actions. This includes Spence (1977), where an incumbent firm faces a prospective entrant and invests in productive capacity during the first period but may choose not to utilize the full capacity during the second period. It also includes Saloner (1987), Admati and Perry (1991), and Romano and Yildirim (2005), where agents can set a lower bound on the action they will choose but retain until the last period the option to choose any action that is at least as large as this lower bound. The present paper belongs to this branch of literature; our contribution is to consider \textit{all} possible commitment structures, that is, all collections of subsets that the leader may choose from in the first period.

A second branch of literature proposes that agents commit to subsets of actions because doing so gives them a strategic advantage. In these models, agents could commit to a single action but typically \textit{choose} to commit to a subset containing more than one action because doing so enables them to credibly threaten potential deviating players. Bade, Haeringer and Renou (2009), Renou (2009), and Dutta and Ishii (2016), embed a strategic-form game into a multi-stage game in which, early on in the game, players can freely restrict their action

\textsuperscript{2}The function $U$ mapping every action of the leader to the payoff she obtains when the follower best-responds plays a central role in our results. A subset of the leader’s action space is called \textit{$U$-monotone} if whenever an action is contained in it so is the upper contour set of that action with respect to $U$. We show that the plausible actions of the leader are $U$-monotone, whereas the simply-plausible ones need not be.
spaces. A player can commit to any action of his choice, but may also choose not to commit at all. This ability to freely choose what to commit to differs from our model and has stark implications: in our setting, if the leader can choose not to commit at all, she can guarantee herself her lowest Cournot payoff, while if she can commit to any action, then she can guarantee herself her Stackelberg payoff. By contrast, we show that if the leader’s ability to commit is limited, she will not only fall short of her Stackelberg payoff but may also fall short of her lowest Cournot payoff. Notwithstanding, Proposition 1 in Bade et al. (2009) constitutes a key building block towards our characterization of the set of simply plausible outcomes. Note that several other papers study what might be construed as a form of limited commitment: some allow agents to pick specific actions but let them revise these choices later on, either at fixed times (Maskin and Tirole, 1988), stochastically (Kamada and Kandori, 2020), or by incurring various costs (Henkel, 2002; Caruana and Einav, 2008).

Finally, our paper belongs to a recent strand of papers that take a base game as given and examine how changing the structure of this game can affect its outcome. For example, Kamenica and Gentzkow (2011), Bergemann and Morris (2016), and Makris and Renou (2023) examine the implications of changing a game’s information structure. Nishihara (1997) and Gallice and Monzón (2019) study instead the effects of changing the order of moves. Salcedo (2017) and Doval and Ely (2020) allow the structure of the game to change in both of these dimensions.

---

3For a multi-stage model of commitment where the base-game itself is an extensive-form game, see Arieli, Babichenko and Tennenholtz (2017).

4Relatedly, Pei (2016) studies a setting where a player can restrict the actions of her opponent. Pei (2016) shows that when a player has limited commitment ability, i.e., cannot reduce her opponent’s action set to a singleton, then it is sometimes strictly optimal not to restrict the opponent’s actions at all.

5For a study of more general commitment devices that include commitments contingent on other players’ commitments, see Kalai, Kalai, Lehrer and Samet (2010).
2 The Model

2.1 Setup

There are two players, a leader and a follower, with action spaces $\mathcal{X} = [x, \bar{x}]$ and $\mathcal{Y} = [y, \bar{y}]$, respectively. A collection $K$ of non-empty subsets of $\mathcal{X}$ covers the leader’s action space.\(^6\) We refer to $K$ as the commitment structure (CST).

There are two periods: in period 1, the leader publicly selects $\mathcal{X}_i \in K$; in period 2, leader and follower simultaneously choose actions $x$ and $y$, with $x$ contained in $\mathcal{X}_i$ and $y$ contained in $\mathcal{Y}$. The resulting payoffs are $u(x, y)$ for the leader and $v(y, x)$ for the follower, where $u$ and $v$ are continuous. We further assume that $u(x, y)$ is strictly quasi-concave in $x$ for all $y \in \mathcal{Y}$, and that $v(y, x)$ is strictly quasi-concave in $y$ for all $x \in \mathcal{X}$. This game is denoted by $G(K)$.

Throughout, our focus is on pure-strategy subgame perfect equilibrium, henceforth simply referred to as equilibrium.

2.2 Definitions and Notation

An action pair $(x, y)$ with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ is referred to as an outcome. We say that an outcome $(x, y)$ is plausible if $(x, y)$ is an equilibrium outcome of $G(K)$ for some commitment structure $K$. An action $x$ is plausible if it is part of a plausible outcome $(x, y)$.

Two salient commitment structures play a central role,

$$K^S := \{\{x\} : x \in \mathcal{X}\} \quad \text{and} \quad K^C := \{\mathcal{X}\};$$

we refer to these as the Stackelberg and Cournot CSTs, respectively. By extension, the equilibrium outcomes of $G(K^S)$ and $G(K^C)$ will be referred to as Stackelberg and Cournot outcomes. The Cournot actions of the leader are the actions of the leader forming part of a Cournot outcome.

A commitment structure $K$ is said to be simple if it partitions the leader’s action space

\(^6\)That is, (i) for all $\mathcal{X}_i \in K$: $\mathcal{X}_i \subseteq \mathcal{X}$; (ii) for all $x \in \mathcal{X}$: $x \in \mathcal{X}_i$ for some $\mathcal{X}_i \in K$. 

6
into intervals. For example, the Stackelberg and Cournot CSTs are simple CSTs. An outcome \((x, y)\) is *simply plausible* if \((x, y)\) is an equilibrium outcome of \(G(K)\), for some simple commitment structure \(K\).

To every action \(x\) of the leader corresponds a unique best response of the follower.\(^7\) We denote this best response by \(R_F(x)\), and let \(U(x)\) be the payoff of the leader from taking action \(x\) when the follower best-responds to \(x\), that is,

\[
U(x) := u(x, R_F(x)).
\]

A subset \(\tilde{X} \subseteq X\) is called *\(U\)-monotone* if \(\tilde{x} \in \tilde{X}\) implies that the upper contour set of \(\tilde{x}\) with respect to \(U\) is contained in \(\tilde{X}\) as well.\(^8\)

### 2.3 Duopoly Example

In this subsection, we illustrate the model in a textbook duopoly setting. Leader and follower are two identical firms, each choosing a quantity in \(X = \mathcal{Y} = \left[0, \frac{2}{2-r}\right]\).\(^9\) A firm producing quantity \(q\) incurs cost \(3q - rq^2/2\), where \(r < 2\) measures the returns to scale. Firms sell at unit price \(4 - (x + y)\). Letting \(u(x, y)\) (respectively, \(v(y, x)\)) be the profit of the leader (respectively, the follower) gives \(v(y, x) = u(y, x)\) and

\[
u(x, y) = x - xy - \left(1 - \frac{r}{2}\right)x^2.
\]

We set for now \(r = 0\). The (unique) Cournot and Stackelberg actions are then, respectively, \(x_C = 1/3\) and \(x_S = 1/2\). Let us consider the simple commitment structure

\[
K = \left\{ \left[0, \frac{3}{5}\right], \left[\frac{3}{5}, 1\right] \right\}.
\]

Figure 1 illustrates this example. Any quantity in the interval \([3/5, 1]\) is such that, whenever the follower best-responds, the leader benefits from deviating to a smaller quantity. Hence,

---

\(^7\)Recall, the follower’s action space is compact, and \(v(y, x)\) is strictly quasi-concave in \(y\).

\(^8\)The upper contour set of an action \(x\) with respect to \(U\) is the set of actions \(\tilde{x}\) such that \(U(\tilde{x}) \geq U(x)\).

\(^9\)Quantities larger than \(2/(2 - r)\) would lead to negative profits no matter what.
any equilibrium of $G(K)$ must be such that the leader produces $3/5$ in the corresponding
subgame. If instead the leader picks $[0, 3/5)$ in the first period, then each firm produces the
Cournot quantity. As $U(3/5) > U(x^C)$, the unique equilibrium is such that in period 1 the
leader chooses the upper interval.

The previous reasoning applies if $3/5$ is replaced by any quantity $x^*$ in the interval
$(1/3, 2/3]$, so all actions in $[1/3, 2/3]$ are simply plausible. We will see below that these
are in fact the only simply-plausible actions: we show in Section 4 that any simply-plausible
outcome is such that the leader’s payoff is bounded by her Stackelberg and Cournot payoffs.
However, we show in Section 5 that the set of plausible outcomes may be larger than the set
of simply-plausible ones; moreover, in general, the Cournot payoffs do not bound the payoffs
that the leader can obtain.

3 Preliminaries

Henceforth, let

$$\eta(\bar{x}, x) := u(\bar{x}, R_F(x)) - u(x, R_F(x)).$$

In words, $\eta(\bar{x}, x)$ measures the leader’s gain from choosing $\bar{x}$ instead of $x$ when the follower
best-responds to $x$. 
Consider an arbitrary commitment structure $K$. Suppose that an equilibrium of $G(K)$ exists. Given $\mathcal{X}_i \in K$, write $\beta(\mathcal{X}_i)$ for the leader’s action in the subgame following $\mathcal{X}_i$. Then $\beta(\mathcal{X}_i) \in \mathcal{X}_i$, and $\eta(x, \beta(\mathcal{X}_i)) \leq 0$ for all $x \in \mathcal{X}_i$. The notion of *admissible pair* summarizes these basic properties.

**Definition 1.** A pair $(K, \beta)$ made up of a commitment structure $K$ and a mapping $\beta : K \to \mathcal{X}$ is said to be admissible if

(a) $\beta(\mathcal{X}_i) \in \mathcal{X}_i$, for all $\mathcal{X}_i \in K$;

(b) $\eta(x, \beta(\mathcal{X}_i)) \leq 0$, for all $x \in \mathcal{X}_i$ and all $\mathcal{X}_i \in K$.

The following characterization of the set of plausible outcomes is immediate.

**Lemma 1.** An outcome $(x, y)$ is plausible if and only if there exist an admissible pair $(K, \beta)$ and $\mathcal{X}_i \in K$, such that

(i) $x = \beta(\mathcal{X}_i)$,

(ii) $U(x) = \max_{\mathcal{X}_j \in K} U(\beta(\mathcal{X}_j))$,

(iii) $y = R_F(x)$.

We say that an admissible pair $(K, \beta)$ implements outcome $(x, y)$ if it satisfies conditions (i)–(iii) of Lemma 1. We can then rephrase the lemma to say that an outcome $(x, y)$ is plausible if and only if some admissible pair $(K, \beta)$ implements it.

## 4 Simple Commitment Structures

This section contains the first part of our analysis; in it, we characterize the set of simply-plausible outcomes. All proofs for this section are in Appendix A.
Denote by $R_L(y)$ the unique best response of the leader to the follower’s action $y$, and define\textsuperscript{10}

$$\phi(x) := R_L(R_F(x)).$$

The fixed points of $\phi$ are thus the Cournot actions of the leader. Let $\mathcal{X}^C$ denote said set of Cournot actions; the notation $x^C_n$ will indicate a generic element of this set.

The following lemma states a central property of admissible pairs $(K, \beta)$ such that $K$ is a CST comprising only intervals. The essence of this lemma is akin to Proposition 1 in Bade, Haeringer and Renou (2007).

**Lemma 2.** Let $K$ be a commitment structure comprising only intervals. Then $(K, \beta)$ is admissible if and only if, for all $\mathcal{X}_i \in K$, one of the following conditions holds:

(i) $\beta(\mathcal{X}_i) \in \mathcal{X}_i \cap \mathcal{X}^C$;

(ii) $\beta(\mathcal{X}_i) = \min \mathcal{X}_i$ and $\phi(\beta(\mathcal{X}_i)) < \beta(\mathcal{X}_i)$;

(iii) $\beta(\mathcal{X}_i) = \max \mathcal{X}_i$ and $\phi(\beta(\mathcal{X}_i)) > \beta(\mathcal{X}_i)$.

The intuition behind the lemma is straightforward. Consider an interval $\mathcal{X}_i$ forming part of a commitment structure $K$, and a mapping $\beta : K \to \mathcal{X}$ such that $\beta(\mathcal{X}_i) \in \mathcal{X}_i$ for all $\mathcal{X}_i \in K$. Suppose that $\phi(\beta(\mathcal{X}_i)) > \beta(\mathcal{X}_i)$. Then, when the follower best-responds to $\beta(\mathcal{X}_i)$, the leader would like to play an action slightly greater than $\beta(\mathcal{X}_i)$. For $(K, \beta)$ to be admissible the leader must therefore be unable to play such a greater action after choosing $\mathcal{X}_i$ in period 1. Since $\mathcal{X}_i$ is an interval, this amounts to saying that $\beta(\mathcal{X}_i)$ is the upper bound of $\mathcal{X}_i$.

Figure 2, panel A, illustrates Lemma 2 in the context of the duopoly example introduced in Subsection 2.3, for $r = 6/5$. The black curve represents the graph of the function $\phi$. The leader’s Cournot actions are $x^C_1 = 0$, $x^C_2 = 5/9$, and $x^C_3 = 5/4$. An admissible pair $(K, \beta)$ is such that every action $\beta(\mathcal{X}_i)$ belonging to a region of the figure with a left-pointing arrow (respectively, right-pointing arrow) is either a Cournot action or the leftmost (respectively, rightmost) element of $\mathcal{X}_i$.

\textsuperscript{10}The leader’s action space is compact and $u(x, y)$ is strictly quasi-concave in $x$. 

10
Our first theorem characterizes the set of simply-plausible outcomes.

**Theorem 1.** An action $x^*$ is simply plausible if and only if the lower contour set of $x^*$ with respect to $U$ contains a Cournot action $x^C_n$ such that

$$
(\phi(x^*) - x^*)(x^C_n - x^*) \geq 0.
$$

(2)

The proof of the \textit{only if} part goes as follows. Pick a simply plausible action $x^*$ such that $\phi(x^*) > x^*$. Reason by contradiction, and assume that all Cournot actions of the leader greater than $x^*$ are in the strict upper contour set of $x^*$. We first show that some Cournot action of the leader has to be greater than $x^*$, and let $x^C_1$ denote the smallest such action. Note that $x^C_1$ cannot be an equilibrium action in the corresponding subgame; something else is, say

\footnote{In particular, it follows that every simply-plausible outcome is an equilibrium outcome of a game $G(K)$ for some simple commitment structure $K$ comprising two elements only.}
We argue that $x_1$ must be greater than $x'_1$ and such that $\phi(x_1) > x_1$. We now repeat the argument above with $x_1$ in place of $x^*$, and so on. We show that the previous recursion must end after finitely many steps, say $n$, for otherwise the fact that the action spaces are compact and the payoff functions are continuous leads to a contradiction. Yet $\phi(x_n) > x_n$ implies $x_n < \pi$, giving another contradiction.

Applying Theorem 1 to the example of Figure 2 shows that the set of simply-plausible actions is equal to $\{0\} \cup [5/17, 5/9] \cup [5/4, 5/2]$ (illustrated in green in panel B of Figure 2). Firstly, Theorem 1 shows that no action in the interval $(0, 5/17)$ is simply plausible, since all of them belong to the strict lower contour set of each Cournot action. Secondly, any $x \in (5/9, 5/4)$ satisfies $\phi(x) > x$ (see panel A). The only Cournot action greater than any of these actions is $x_3^C$. As $U(x_3^C) > U(x)$ for all $x \in (5/9, 5/4)$, we conclude using Theorem 1 that no action in this interval is simply plausible. Mirror arguments show that all actions in $\{0\} \cup [5/17, 5/9] \cup [5/4, 5/2]$ are simply plausible.\footnote{An action $x^* \in [5/17, 5/9] \cup [5/4, 5/2]$ is implemented, for instance, by the pair $(K, \beta)$ where $K = \{[0, x^*), [x^*, 5/2]\}$, $\beta([0, x^*)) = 0$, and $\beta([x^*, 5/2]) = x^*$.}

By construction, the leader’s Stackelberg payoff provides an upper bound for the payoffs attainable by the leader under any CST. Theorem 1 shows that the Cournot payoffs provide a corresponding lower bound for simple CSTs. Moreover, Theorem 1 implies that if an action is in the upper contour set of all Cournot actions, then that action must be simply plausible. The following corollary records these observations.

**Corollary 1.** All simply-plausible actions belong to the upper contour set of a Cournot action with respect to $U$. Furthermore, any action in the intersection of these upper contour sets is simply plausible.

## 5 Beyond Simple Commitment Structures

We saw in the previous section that all simply-plausible outcomes guarantee the leader at least her lowest Cournot payoff. The following examples show that the conditions imposed on
simple CSTs are crucial: in general, an outcome may be plausible and give the leader a payoff that is less than her minimum Cournot payoff.

**Example A.** Consider the duopoly example introduced in Subsection 2.3, with \( r = 4/5 \) and commitment structure

\[
K = \left\{ \left[ \frac{1}{8}, \frac{1}{3} \right], \left[ 0, \frac{1}{8} \right] \cup \left( \frac{1}{3}, \frac{5}{3} \right) \right\}.
\]

As one of its elements is not an interval, \( K \) is not a simple CST. Figure 3 illustrates this example. The subgame following the leader’s choice of \((1/8, 1/3)\) possesses a unique equilibrium, in which the leader produces 1/3. The other subgame has two equilibria: one yielding the Cournot outcome, \( x^C = 5/11 \), the other involving the leader choosing quantity 1/8. As \( U(1/8) < U(1/3) < U(x^C) \), we see that \( G(K) \) possesses two equilibria: one in which the leader produces 1/3, and one in which the leader produces \( x^C \). In the former equilibrium, the leader anticipates that if she were to select \([0, 1/8] \cup (1/3, 5/3)\) in period 1, the follower would respond by producing a quantity larger than \( x^C \). Consequently, the leader settles for the quantity 1/3, and obtains less than the Cournot payoff \( U(x^C) \).

**Example B.** Consider the following setting. The action spaces are \( \mathcal{X} = \mathcal{Y} = [0, 1] \). The payoffs of the leader are given by

\[
u(x, y) = xy + (1-x)(1-y) - \frac{1}{2} \left( x - \frac{1}{2} \right)^2 - \frac{3}{2} \left( y - \frac{1}{2} \right)^2 ;
\]
the payoffs of the follower are given by $v(y, x) = u(y, x)$. In this setting, the leader’s Stackelberg actions are 0, 1/2, and 1, and these are also the leader’s Cournot actions. The commitment structure is

$$K = \{[0, x^*], [1 - x^*, 1]\},$$

where $x^*$ denotes some action in $(1/2, 1)$. As $x^* > 1 - x^*$, $K$ does not partition the leader’s action space; so $K$ is not a simple CST. Figure 4 illustrates this example. The subgame induced by the leader’s choice of $[0, x^*]$ has an equilibrium in which the leader chooses the action $x^*$. Symmetrically, the subgame induced by the leader’s choice of $[1 - x^*, 1]$ has an equilibrium in which the leader chooses the action $1 - x^*$. Therefore, since $U(x^*) = U(1 - x^*)$, both $x^*$ and $1 - x^*$ are plausible. In both cases, the leader obtains less than her Cournot payoff $U(1/2)$.

A simple CST has two defining properties: it partitions the leader’s action space and consists solely of intervals. The previous examples illustrate that relaxing either of these conditions can expand the set of outcomes that are plausible. In particular, relaxing either of these conditions can induce the leader to obtain less than her minimum Cournot payoff. Subsection 5.1 characterizes the set of outcomes induced by CSTs comprising only intervals. CSTs comprising non-convex sets are examined in Subsection 5.2: we characterize the set of outcomes induced by CSTs which partition the leader’s action space and show that every
plausible outcome is plausible under such a CST.

5.1 Commitment Structures Comprising Only Intervals

We say that an outcome \((x, y)\) is \(I\)-plausible if it is an equilibrium outcome of \(G(K)\) for some commitment structure \(K\) comprising only intervals. Every simply-plausible outcome is \(I\)-plausible. However, as Example B illustrated, an outcome may be \(I\)-plausible even though it is not simply plausible. The following theorem characterizes the set of \(I\)-plausible outcomes.\(^\text{13}\)

All proofs for this subsection are in Appendix B.

**Theorem 2.** The set of \(I\)-plausible actions is \(U\)-monotone. An action \(x^*\) is \(I\)-plausible if and only if the lower contour set of \(x^*\) with respect to \(U\) includes actions \(x'\) and \(x''\) such that

\[
\phi(x') \leq x' \leq x'' \leq \phi(x'').
\]

\(^3\)

Suppose that \(x\) is \(I\)-plausible, and let \((K, \beta)\) be an admissible pair that implements \(x\) in which the commitment structure \(K\) contains only intervals. Pick \(x'\) in the upper contour set of \(x\) with respect to \(U\), and consider \(K' := K \cup \{x'\}\), and \(\beta' : K' \to \mathcal{X}\) such that \(\beta'(X_i) = \beta(X_i)\) for all \(X_i \in K\) and \(\beta'(\{x'\}) = x'\). The commitment structure \(K'\) contains only intervals. Furthermore, the pair \((K', \beta')\) is evidently admissible and implements \(x'\). We conclude that the \(I\)-plausible actions are \(U\)-monotone.

The if part of the theorem is straightforward. Let \(x^*\) be such that the lower contour set of \(x^*\) with respect to \(U\) includes actions \(x'\) and \(x''\) satisfying (3). Consider \(K = \{\{x^*\}, [x, x'], [x', x]\}\), and \(\beta\) given by \(\beta(\{x^*\}) = x^*, \beta([x, x']) = x'', \beta([x', x]) = x'\). By Lemma 2, the pair \((K, \beta)\) is admissible; it implements \(x^*\) since \(U(x^*) \geq \max\{U(x'), U(x'')\}\).\(^4\)

\(^{13}\)Note that in case \(x^*\) is simply plausible, Theorem 1 tells us that the set of Cournot actions in the lower contour set of \(x^*\) is non-empty; let \(x_n^C\) be such an action. As \(x_n^C = \phi(x_n^C)\) choosing \(x' = x'' = x_n^C\) immediately provides two actions satisfying the conditions of Theorem 2.

\(^{14}\)In particular, it follows that every \(I\)-plausible outcome is an equilibrium outcome of a game \(G(K)\) for some commitment structure \(K\) comprising at most three elements.
The proof of the only if part of the theorem first establishes that if \( x^* \) is I-plausible then some action \( x \) in the lower contour set of \( x^* \) must satisfy \( \phi(x) \leq x \), and some action \( x \) in the lower contour set of \( x^* \) must satisfy \( \phi(x) \geq x \). Next, suppose that every action \( x \) in the lower contour set of \( x^* \) for which \( \phi(x) \geq x \) is strictly smaller than every action \( x \) in the lower contour set of \( x^* \) for which \( \phi(x) \leq x \), and pick \( x^\dagger \) in between these two subsets of actions. Now let \((K, \beta)\) be an admissible pair such that \( K \) is a CST comprising only intervals. By Lemma 2, if \( \mathcal{X}_i \in K \) is an interval containing \( x^\dagger \) then either \( \beta(\mathcal{X}_i) \geq x^\dagger \) and \( \phi(\beta(\mathcal{X}_i)) \geq \beta(\mathcal{X}_i) \), or \( \beta(\mathcal{X}_i) \leq x^\dagger \) and \( \phi(\beta(\mathcal{X}_i)) \leq \beta(\mathcal{X}_i) \). Yet every action \( x \) in the lower contour set of \( x^* \) for which \( \phi(x) \geq x \) is strictly smaller than \( x^\dagger \), while every action \( x \) in the lower contour set of \( x^* \) for which \( \phi(x) \leq x \) is strictly greater than \( x^\dagger \). We conclude that \( \beta(\mathcal{X}_i) \) must belong to the strict upper contour set of \( x^* \). So \((K, \beta)\) cannot implement \( x^* \).

Note that if a Cournot action \( x^C_n \) belongs to the lower contour set of an action \( x^* \), then setting \( x' = x'' = x^C_n \) in Theorem 2 proves that \( x^* \) is I-plausible. We thus obtain the following corollary:

**Corollary 2.** All actions in the upper contour set of a Cournot action with respect to \( U \) are I-plausible.

Applying Theorem 2 to the example of Figure 2 shows that the I-plausible actions are \( \{0\} \cup [5/17, 5/2] \). In particular, actions in the interval \((5/9, 5/4)\) are I-plausible but are not simply plausible.\(^{15}\)

The previous analysis has shown that an outcome may be I-plausible even though it is not simply plausible and that I-plausible outcomes can give the leader a smaller payoff than her lowest Cournot payoff. However, in certain prominent cases, every I-plausible outcome does guarantee the leader at least her minimum Cournot payoff. For example, consider a setting with a unique Cournot action \( x^C \). We have in this case \( \{x : \phi(x) \geq x\} = [x, x^C] \)

\(^{15}\)An action \( x^* \in (5/9, 5/4) \) is for instance implemented by the pair \((K, \beta)\) where \( K = \{\mathcal{X}, [0, x^*]\}, \beta(\mathcal{X}) = 0, \) and \( \beta([0, x^*]) = x^* \). To see that no \( x^* \in (0, 5/17) \) is I-plausible, notice that the intersection between \( \{x : \phi(x) \geq x\} \) and the lower contour set of \( x^* \) with respect to \( U \) is empty.
and \( \{x: \phi(x) \leq x\} = [x^C, \pi] \). Applying Theorem 2 thus shows that all I-plausible actions must belong to the upper contour set of \( x^C \) with respect to \( U \). As all actions in the upper contour set of \( x^C \) are simply plausible (Corollary 1), we conclude that in this case an action is I-plausible if and only if it is in the upper contour set of \( x^C \) with respect to \( U \). A similar result holds when \( U \) is either quasi-convex or quasi-concave.

**Proposition 1.** If there exists a unique Cournot outcome, or if \( U \) is either quasi-convex or quasi-concave, then an action is I-plausible if and only if it belongs to the union of the upper contour sets of the Cournot actions with respect to \( U \).

### 5.2 Partitional Commitment Structures

We say that an outcome \((x, y)\) is \( P\)-plausible if it is an equilibrium outcome of \( G(K) \) for some commitment structure \( K \) which partitions the leader’s action space. Every simply-plausible outcome is \( P\)-plausible. However, as Example A illustrated, an outcome may be \( P\)-plausible even though it is not simply plausible. We start this subsection by establishing that any plausible outcome is, in fact, \( P\)-plausible.\(^{16}\) All proofs for this section are in Appendix C.

**Proposition 2.** An outcome is plausible if and only if it is \( P\)-plausible. Moreover, the set of \( P\)-plausible outcomes is \( U\)-monotone.

The \( if \) part of the proposition is trivial. The gist of the proof of the \( only if \) part is as follows. Let \((x, y)\) be plausible, and \((K, \beta)\) be an admissible pair that implements \((x, y)\). Suppose that \( K = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\} \), and\(^{17}\)

\[
\beta(\mathcal{X}_i) \notin \bigcup_{j \neq i} \mathcal{X}_j, \quad \text{for } i = 1, \ldots, n. \tag{4}
\]

Now let \( \mathcal{X}'_i = \mathcal{X}_i \setminus \bigcup_{j < i} \mathcal{X}_j \), for \( i = 1, \ldots, n \), and \( K' = \{\mathcal{X}'_1, \ldots, \mathcal{X}'_n\} \). Then \( K' \) partitions \( \mathcal{X} \) and (4) implies \( \beta(\mathcal{X}'_i) \in \mathcal{X}'_i \), for \( i = 1, \ldots, n \). Letting \( \beta'(\mathcal{X}'_i) = \beta(\mathcal{X}_i) \) for \( i = 1, \ldots, n \), we

\(^{16}\)We thank an anonymous referee for pointing this out to us.

\(^{17}\)See the Appendix C for the general case.
see that \((K', \beta')\) is admissible, since \(X'_i \subseteq X_i\). Finally, \((K', \beta')\) evidently implements \((x, y)\), so \((x, y)\) is P-plausible.

Since any plausible outcome is also P-plausible, the set of P-plausible outcomes is typically very large. To make progress, we restrict attention in the rest of this subsection to settings that satisfy the following three regularity conditions:

\(\text{(RC1)}\) there exists a unique and interior Cournot outcome;

\(\text{(RC2)}\) all externalities are either strictly positive or strictly negative;\(^{18}\)

\(\text{(RC3)}\) payoffs are either strictly supermodular or strictly submodular.\(^{19}\)

For instance, the duopoly example of Subsection 2.3 satisfies all three conditions as long as the returns to scale are not too large \((r < 1)\). When \(u\) and \(v\) are twice differentiable, conditions (RC2) and (RC3) become \(u_2 v_2 > 0\) and \(u_{12} v_{12} > 0\), respectively. Slightly abusing notation, whether or not the payoffs are differentiable, \(u_2 > 0\) will indicate positive externalities, and \(u_2 < 0\) negative externalities. Similarly, \(u_{12} > 0\) will indicate supermodular payoffs, and \(u_{12} < 0\) submodular payoffs.

As \(u(\cdot, R_F(x))\) is strictly quasi-concave, for \(x \leq x^C\) the set of actions which the leader would weakly prefer to \(x\) when the follower best-responds to \(x\) can be written as \([x, \gamma(x)]\), for some \(\gamma(x)\).\(^{20}\) Similarly, for \(x \geq x^C\) said set of actions can be written as \([\gamma(x), x]\).\(^{21}\) This

\(18\)Formally, externalities are strictly positive (respectively, negative) if \(u\) and \(v\) are strictly increasing (respectively, decreasing) in their second arguments.

\(19\)Formally, \(u\) (and similarly \(v\)) is strictly supermodular (respectively, submodular) if \(u(x', y') + u(x, y) > u(x', y) + u(x, y')\) (respectively, \(u(x', y') + u(x, y) < u(x', y) + u(x, y')\)) for all \(x' > x\) and \(y' > y\). Supermodular payoffs capture strategic complementarities; submodular payoffs capture strategic substitutabilities.

\(20\)If \(x = x^C\) then \(\gamma(x) = x^C\). Otherwise, either \(u(\pi, R_F(x)) > u(x, R_F(x))\) in which case \(\gamma(x) = \pi\), or \(u(\pi, R_F(x)) \leq u(x, R_F(x))\) in which case \(\gamma(x)\) is the unique action different from \(x\) making the leader indifferent between choosing \(x\) and \(\gamma(x)\) when the follower best-responds to \(x\).

\(21\)If \(x = x^C\) then \(\gamma(x) = x^C\). Otherwise, either \(u(x, R_F(x)) > u(x, R_F(x))\) in which case \(\gamma(x) = x\), or \(u(x, R_F(x)) \leq u(x, R_F(x))\) in which case \(\gamma(x)\) is the unique action different from \(x\) making the leader indifferent between choosing \(x\) and \(\gamma(x)\) when the follower best-responds to \(x\).
defines a mapping $\gamma : X \to X$; note that, as $u$ and $R_F$ are continuous, so is $\gamma$. Finally, let

$$S := \begin{cases} 
\{ x : x \leq \gamma(x) \leq x^C \} & \text{if } u_2u_{12} > 0, \\
\{ x : x^C \leq \gamma(x) \leq x \} & \text{if } u_2u_{12} < 0.
\end{cases}$$

Since $\gamma$ is continuous, the set $S$ is compact. Moreover, this set evidently contains $x^C$. We can now characterize the set of plausible outcomes.

**Theorem 3.** Suppose (RC1)–(RC3) hold. The set of plausible actions coincides with the upper level set of $U := \min_{x \in S} U(\gamma(x))$ with respect to $U$.\(^{22}\)

Our proof that any $x$ such that $U(x) \geq U$ is a plausible action rests on the following methodology.\(^{23}\) We first collect all $x'$ in the strict upper contour set of $x$ with respect to $U$, and place these $x'$ into a set together with some worse alternative, say $x''$. Then, provided $(x'', R_F(x''))$ is a Nash equilibrium of the corresponding subgame, the action $x$ we started from can readily be made into an equilibrium outcome.

To be more specific, consider the case in which $u_2 > 0$ and $u_{12} > 0$, that is, players’ actions have positive externalities and are strategic complements. In this case, $S = \{ x : x \leq \gamma(x) \leq x^C \}$, and the function $U$ is increasing in the range $[\hat{x}, x^C]$.\(^{24}\) Now pick some action $\hat{x} \in S$ and let $x^*$ be in the upper contour set of $\gamma(\hat{x})$ with respect to $U$; we will argue that $x^*$ is plausible. The case $\hat{x} = x^C$ being trivial, suppose $\hat{x} \neq x^C$ and so $\hat{x} < \gamma(\hat{x}) < x^C$.

Let $X_1 = [\hat{x}, x^C] \cup (\gamma(\hat{x}), x^*) \cup (x^*, \bar{x}]$. Next, let $K$ be the partition of the leader’s action space containing $X_1$, the interval $(\hat{x}, \gamma(\hat{x})]$ and the singleton $\{ x^* \}$. Lastly, let $\beta(X_1) = \hat{x}$, $\beta((\hat{x}, \gamma(\hat{x}])) = \gamma(\hat{x})$ and $\beta(\{ x^* \}) = x^*$. We claim that the pair $(K, \beta)$ is admissible. By definition of $\gamma$, the set of actions which the leader weakly prefers to $\hat{x}$ when the follower

\(^{22}\)The upper level set of $U$ with respect to $U$ is defined as $\{ x : U(x) \geq U \}$.

\(^{23}\)The proof that the upper level set of $U$ includes the set of plausible actions is more technical. We refer the interested reader to the appendix.

\(^{24}\)Intuitively, increasing the action of the leader induces the follower to increase her action too (due to strategic complementarities), and this benefits the leader (since externalities are positive).
best-responds to \( \hat{x} \) can be written as \([\hat{x}, \gamma(\hat{x})]\). So \( \eta(x, \hat{x}) \leq 0 \) for all \( x \in \mathcal{X}_1 \). Furthermore as \( \gamma(\hat{x}) < x^C \) then \( \phi(\gamma(\hat{x})) > 0 \) and therefore \( \eta(x, \gamma(\hat{x})) \leq 0 \) for all \( x \in (\hat{x}, \gamma(\hat{x})] \). Finally, we claim that \((K, \beta)\) implements \( x^* \). Indeed, \( U(\hat{x}) < U(\gamma(\hat{x})) \leq U(x^*) \); the first inequality follows from \( \hat{x} < \gamma(\hat{x}) \) and the fact that \( U \) is increasing over \([x, \gamma(\hat{x})]\); the second inequality is immediate, since \( x^* \) is in the upper contour set of \( \gamma(\hat{x}) \) by definition.\(^{25}\)

Figure 5 illustrates Theorem 3 in the context of the duopoly example from Subsection 2.3 with \( r = 4/5 \). In panel A, the black curve represents the graph of the function \( \phi \), which crosses the 45-degree line at \( x^C = 5/11 \). In this example \( u_2 < 0 \) and \( u_{12} < 0 \), so \( S = \{x : x \leq \gamma(x) \leq x^C\} \). The gray curve represents the graph of the function \( \gamma \): we see that \( S = [0, x^C] \) and \( \gamma(S) = [5/18, x^C] \). Panel B depicts the graph of the function \( U \). Minimizing \( U(\gamma(x)) \) over \( S \) shows that \( \underline{U} = U(\gamma(0)) = U(5/18) \). So the set of plausible actions is \([5/18, x^t]\), where \( U(x^t) = U(5/18) \). Note that \( \underline{U} \) is less than the leader’s Cournot payoff \( U(x^C) \), hence some plausible outcomes give the leader less than her Cournot payoff.

\(^{25}\)This construction immediately implies that all plausible outcomes are an equilibrium outcome of \( G(K) \) for some commitment structure \( K \) with at most three elements.
that $\gamma'(x^C) > 0$ if and only if $R'_L(y^C)R'_F(x^C) > 1/2$. In other words, whenever the best-response functions are sufficiently steep, some plausible outcomes give the leader less than her Cournot payoff.

**Proposition 3.** Suppose $u$ and $v$ are twice differentiable and (RC1)–(RC3) hold. If $R'_L(y^C)R'_F(x^C) > 1/2$ then $U < U(x^C)$.

We have seen that the set of plausible outcomes may be larger than the set of simply-plausible ones. Proposition 2 shows that all plausible outcomes can be generated by partitional CSTs. A natural question is whether an even more restrictive class of CSTs can generate the entire set of plausible outcomes. The answer is yes, and one such class consists of what we call *quasi-simple* CSTs: a commitment structure is quasi-simple if it partitions the leader’s action space and contains at most one element that is not an interval. We say that an outcome is *quasi-simply plausible* if it is an equilibrium outcome of $G(K)$ for some quasi-simple commitment structure $K$.

**Proposition 4.** Suppose (RC1)–(RC3) hold. Then every plausible outcome is quasi-simply plausible.

The proof follows directly from the construction in the proof of Theorem 3. In fact, it can be shown that if $U$ is either quasi-concave or quasi-convex then all plausible outcomes can be implemented with a partition of the leader’s action space such that each partition element is either an interval or a union of two intervals. Effectively, these commitment structures are such that the leader plainly commits to choosing an action: (a) inside or outside an interval, (b) below or above a cutoff.

### 5.3 Equilibrium Refinements

We saw that the Stackelberg and Cournot payoffs provide the bounds of the payoffs attainable by the leader under any simple CST. A natural question is whether some equilibrium refinement ensures that the Stackelberg and Cournot payoffs provide the bounds of the payoffs attainable by the leader under arbitrary CSTs.
Forward induction type of arguments eliminate some, but not all, equilibria giving the leader less than her Cournot payoffs. For instance, consider the setting of Example B at the beginning of this section, but this time with commitment structure  

\[
\{ \left[ \frac{1}{9}, \frac{4}{9} \right], \left[ 0, \frac{1}{9} \right] \cup \left[ \frac{4}{9}, 1 \right] \}.
\]

Figure 6 depicts the graph of \( U(x) \). The subgame induced by the leader’s choice of \( \left[ \frac{1}{9}, \frac{4}{9} \right) \) has a unique equilibrium, in which the leader chooses the action 1/9. The other subgame has an equilibrium in which the leader picks 4/9. As \( U(4/9) > U(1/9) \), an equilibrium exists in which the leader chooses 4/9. Yet, \( U(4/9) < U(x^n) \), so the leader obtains a payoff smaller than her Cournot payoff. Since the subgame off the equilibrium path possesses a unique equilibrium, forward induction type of arguments have no bite.

One alternative is to restrict attention to equilibria that select, in every period-2 subgame, the best continuation equilibrium from the perspective of the leader. In this case, any subgame induced by the leader’s period-1 choice of a subset containing a Cournot action must give the leader a payoff at least as large as that Cournot payoff. Consequently, any such equilibrium ensures that the leader obtains at least her maximum Cournot payoff.

\[26\text{See Myerson (1997) for a discussion of the merits and flaws of forward induction.}\]
We now apply our results to the duopoly setting introduced in Subsection 2.3. Figure 7 shows the action set $X = [0, 2/(2 - r)]$ as well as the Cournot and Stackelberg actions of the leader. For $r < 1$, the unique Cournot outcome is such that leader and follower produce the same amount. For $r > 1$ the three Cournot outcomes are, respectively, such that only the leader produces a positive amount, leader and follower produce the same amount, or only the follower produces a positive amount. Whenever $r < r^S$ the (unique) Stackelberg outcome is such that both firms produce a positive amount, whereas if $r > r^S$ only the leader produces a positive amount.\footnote{For $r \in (r^S, 1)$, the Stackelberg action is larger than the quantity a monopolist would choose; the two are the same for $r \geq 1.$}

The green area in Figure 7 indicates the set of simply plausible actions. For $r < 1$, the green area corresponds to the upper contour set with respect to $U$ of the unique Cournot action (Corollary 1); for $r > 1$, the characterization of the green area is illustrated in the example of Figure 2. Every action $x$ in the green area is part of the equilibrium outcome for the simple commitment structure

$$K = \left\{ [0, x), \left[ x, \frac{2}{2 - r} \right] \right\}.$$
The red area in Figure 7 indicates the set of actions that are plausible but not simply plausible. For $r < 1$, the characterization of this area relies on Theorem 3 and the observation that $\gamma(0) < x^C \iff r > 2 - \sqrt{2}$; for $r > 1$, Theorem 2 implies that any action in between the Stackelberg action and the second-highest Cournot action is plausible. Any action in the red area is part of an equilibrium outcome for a quasi-simple CST.

Next, we consider outcomes that maximize some familiar objective functions within the set of plausible outcomes. Formally, for a given objective function $W$, we solve the maximization problem:

$$\max W(x, R_F(x)) \quad \text{s.t. } x \text{ is plausible.}$$

We first examine the plausible outcomes associated with the largest attainable profit for one of the players (i.e., where $W = u$ or $W = v$). The Stackelberg outcome is plainly the best plausible outcome for the leader. On the other hand, since $v_2$ is here negative, the optimal plausible outcome for the follower involves the leader producing as little as plausibly possible. The proposition which follows summarizes these observations. In the rest of this section, $\underline{x}$ (respectively, $\overline{x}$) indicates the smallest (respectively, largest) plausible action of the leader.

**Proposition 5.** Suppose $u$ is given by (1) and $v(y, x) = u(y, x)$. If $W = u$, the unique solution of (DP) is $x^S$. If $W = v$, the unique solution of (DP) is $\underline{x}$.

The Stackelberg CST is optimal for the leader. The Cournot CST is optimal for the follower if and only if $r \notin (2 - \sqrt{2}, 1)$.

Next, we examine outcomes that maximize either consumer surplus, producer surplus, or total welfare (i.e., the sum of producer and consumer surplus). Consumer surplus and

---

28 Any action $x$ in the white area is not plausible as $u(x, R_F(x)) < 0$, and the leader can ensure zero profit with action $x = 0$.

29 For $r < 1$, any action $x$ in the lower red area ($x < x^C$) is part of an equilibrium outcome for the commitment structure $K = \{(0, x), \{0\} \cup (x, 2/(2 - r))\}$, while any $x$ in the upper red area ($x > x^C$) for $K = \{(0, \gamma(0)), \{0\} \cup (\gamma(0), x), [x, 2/(2 - r)]\}$. For $r > 1$, any $x$ in the red area is part of an equilibrium outcome for $K = \{\{x\}, [0, x) \cup (x, 2/(2 - r))\}$. 

---
producer surplus are defined as
\[ CS(x, y) = \frac{(x + y)^2}{2}, \quad PS(x, y) = u(x, y) + v(y, x). \]

**Proposition 6.** Suppose \( u \) is given by (1) and \( v(y, x) = u(y, x) \).

(i) If \( W = CS \), the unique solution of (DP) is \( \overline{x} \).

(ii) If \( W = PS \), the unique solution of (DP) is \( x^C \) if \( r < r^\dagger \), and \( x^S \) if \( r^\dagger < r < 1 \);\(^{30}\) if \( r > 1 \) then the solutions are \( x^C_3 \) and 0.\(^{31}\)

(iii) If \( W = CS + PS \), the unique solution of (DP) is \( \overline{x} \).

Part (i) of Proposition 6 is explained as follows.\(^{32}\) Firstly, we show that consumer surplus is a convex function of the quantity which the leader produces. The problem therefore reduces to choosing between \( \overline{x} \) and \( \overline{\overline{x}} \). Inducing the leader to produce \( \overline{\overline{x}} \) instead of \( \overline{x} \) is optimal as it exploits the strategic motive to produce large quantities.

Part (ii) of Proposition 6 is straightforward. With decreasing returns to scale, producer surplus is maximized by inducing both firms to produce the same quantity; in this case, the Cournot CST is producer-optimal. By contrast, with large returns to scale, producer surplus is maximized by letting one firm acquire a bigger market share than the other. In particular, for very large returns to scale, producer surplus is maximized when one firm acts as a monopolist. Consequently, the Cournot CST is producer-optimal for extreme returns to scale, whereas the Stackelberg CST is producer-optimal for sufficiently large returns to scale.

Part (iii) of Proposition 6 follows from the fact that producer surplus tends to be less sensitive than consumer surplus to the quantity that the leader produces. Thus, maximizing total welfare implies maximizing consumer surplus.

\[^{30}r^\dagger := 2 - \frac{\sqrt{3(9-\sqrt{78})}}{3} - \frac{1}{\sqrt{3(9-\sqrt{78})}}.\]

\[^{31}x^C_3 = \frac{1}{2-r} \text{ is the highest of the three Cournot equilibria.}\]

\[^{32}\text{The proof of Proposition 6 is available upon request.}\]
7 Conclusion

The Stackelberg leadership model assumes that the leader can commit to any action she might choose. Our paper takes a different view: we only assume that the leader can commit not to take certain subsets of actions.

We provide a tractable model of commitment that encompasses the Stackelberg and Cournot models as special cases but also enables us to capture situations of limited commitment. We characterize the set of outcomes resulting from all possible commitment structures, and shed light thereby on the “limits of limited commitment”. Our results highlight that, more than commitment, what matters is the precise form that commitment takes. For instance, we show that whereas the Stackelberg and Cournot payoffs provide the bounds of the payoffs attainable by the leader under some appropriately defined class of “simple” commitment structures, this property fails to hold more generally.

References

Admati, A. R. and Perry, M. (1991) Joint Projects without Commitment, Review of Economic Studies, 58, 259–276.

Arieli, I., Babichenko, Y. and Tennenholtz, M. (2017) Sequential Commitment Games, Games and Economic Behavior, 105, 297–315.

Bade, S., Haeringer, G. and Renou, L. (2007) More Strategies, More Nash Equilibria, Journal of Economic Theory, 135, 551–557.

Bade, S., Haeringer, G. and Renou, L. (2009) Bilateral Commitment, Journal of Economic Theory, 144, 1817–1831.

Barro, R. J. and Gordon, D. B. (1983) Rules, Discretion and Reputation in a Model of Monetary Policy, Journal of Monetary Economics, 12, 101–121.
Bergemann, D. and Morris, S. (2016) Bayes Correlated Equilibrium and the Comparison of Information Structures in Games, *Theoretical Economics*, **11**, 487–522.

Caruana, G. and Einav, L. (2008) Production Targets, *RAND Journal of Economics*, **39**, 990–1017.

Doval, L. and Ely, J. C. (2020) Sequential Information Design, *Econometrica*, **88**, 2575–2608.

Dutta, R. and Ishii, R. (2016) Dynamic Commitment Games, Efficiency and Coordination, *Journal of Economic Theory*, **163**, 699–727.

Gallice, A. and Monzón, I. (2019) Co-operation in Social Dilemmas Through Position Uncertainty, *Economic Journal*, **129**, 2137–2154.

Henkel, J. (2002) The 1.5th Mover Advantage, *RAND Journal of Economics*, **33**, 156–170.

Kalai, A. T., Kalai, E., Lehrer, E. and Samet, D. (2010) A Commitment Folk Theorem, *Games and Economic Behavior*, **69**, 127–137.

Kamada, Y. and Kandori, M. (2020) Revision Games, *Econometrica*, **88**, 1599–1630.

Kamenica, E. and Gentzkow, M. (2011) Bayesian Persuasion, *American Economic Review*, **101**, 2590–2615.

Kydland, F. E. and Prescott, E. C. (1977) Rules Rather than Discretion: The Inconsistency of Optimal Plans, *Journal of Political Economy*, **85**, 473–492.

Makris, M. and Renou, L. (2023) Information Design in Multi-stage Games, *Theoretical Economics*, **18**, 1475–1509.

Maskin, E. and Tirole, J. (1988) A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs, *Econometrica*, **56**, 549–569.

Myerson, R. B. (1997) *Game Theory: Analysis of Conflict*, Harvard University Press.
Nishihara, K. (1997) A Resolution of N-Person Prisoners’ Dilemma, *Economic Theory*, 10, 531–540.

Pei, H. D. (2016) When Does Restricting Your Opponent’s Freedom Hurt You?, *Games and Economic Behavior*, 100, 234–239.

Renou, L. (2009) Commitment Games, *Games and Economic Behavior*, 66, 488–505.

Romano, R. and Yildirim, H. (2005) On the Endogeneity of Cournot-Nash and Stackelberg Equilibria: Games of Accumulation, *Journal of Economic Theory*, 120, 73–107.

Salcedo, B. (2017) Implementation Without Commitment in Moral Hazard Environments, manuscript.

Saloner, G. (1987) Cournot Duopoly With Two Production Periods, *Journal of Economic Theory*, 42, 183–187.

Spence, A. M. (1977) Entry, Capacity, Investment and Oligopolistic Pricing, *Bell Journal of Economics*, 8, 534–544.

Tirole, J. (1988) *The Theory of Industrial Organization*, The MIT Press, 1st edn.
A Appendix of Section 4

Proof of Lemma 2: We prove the only if part of the lemma; the proof of the other part is similar. Suppose that \((K, \beta)\) constitutes an admissible pair. Reason by contradiction, and suppose that we can find \(X_i \in K\) such that \(\phi(\beta(X_i)) < \beta(X_i)\) while \(\beta(X_i) \neq \min X_i\). The function \(\eta(.; \beta(X_i))\) is strictly quasi-concave, maximized at \(\phi(\beta(X_i))\), and satisfies \(\eta(\beta(X_i), \beta(X_i)) = 0\). So \(\eta(x, \beta(X_i)) > 0\) for all \(x \in [\phi(\beta(X_i)), \beta(X_i)]\). Since \(X_i\) is an interval, \(\beta(X_i) \in X_i\), and \(\beta(X_i) \neq \min X_i\), we can find \(\varepsilon > 0\) such that \((\beta(X_i) - \varepsilon, \beta(X_i)) \subset X_i\). Coupling the previous remarks shows the existence of \(x \in X_i\) such that \(\eta(x, \beta(X_i)) > 0\); this contradicts the assumption that \((K, \beta)\) is admissible. Hence, \(\phi(\beta(X_i)) < \beta(X_i)\) implies \(\beta(X_i) = \min X_i\). Analogous arguments show that \(\phi(\beta(X_i)) > \beta(X_i)\) implies \(\beta(X_i) = \max X_i\). \(\blacksquare\)

Proof of Theorem 1: The if part of the theorem was proven in the text; we prove here the converse. In the rest of the appendix, the upper contour set of \(x\) with respect to \(U\) will be denoted by \(Q_{\geq}(x)\), that is,

\[
Q_{\geq}(x) := \{ \tilde{x} : U(\tilde{x}) \geq U(x) \}.
\]

The sets \(Q_{<}(x)\), \(Q_{\leq}(x)\), and \(Q_{>}(x)\) are similarly defined.

Pick an arbitrary simply-plausible action \(x^*\). We aim to prove the existence of a Cournot action \(x^*_n \in Q_{\leq}(x^*)\) such that (2) holds. If \(\phi(x^*) = x^*\), just take \(x^*_n = x^*\); we treat below the case in which \(\phi(x^*) > x^*\) (the remaining case is analogous). Reason by contradiction, and suppose that

\[
\mathcal{X}^C \cap (x^*, \overline{x}] \cap Q_{\leq}(x^*) = \emptyset. \tag{5}
\]

Let \((K, \beta)\) be an admissible pair that implements \(x^*\), with \(K\) a simple CST. We will show that \(K\) cannot be finite. By Berge’s maximum theorem, both \(R_F\) and \(R_L\) are continuous, thus \(\phi\) is continuous as well. As \(\phi(x^*) > x^*\) and \(\phi(\overline{x}) \leq \overline{x}\), the intermediate value theorem shows that

\[
\mathcal{X}^C \cap (x^*, \overline{x}] \neq \emptyset.
\]

Note that the continuity of the function \(\phi\) implies the compactness of \(\mathcal{X}^C\). So \(\mathcal{X}^C \cap (x^*, \overline{x}] = \mathcal{X}^C \cap [x^*, \overline{x}]\) possesses a smallest element, that we denote by \(x^*_1\). Let \(X_1\) be the member of \(K\)
containing \( x_1^C \). Then Lemma 2 combined with (5) gives

\[
\beta(\mathcal{X}_1) \in (x_1^C, \bar{x}] \cap \{x : \phi(x) > x\}.
\]

Now let \( x_2^C \) be the smallest Cournot action greater than \( \beta(\mathcal{X}_1) \), and denote by \( \mathcal{X}_2 \) the member of \( K \) containing \( x_2^C \). The same logic as above gives \( \beta(\mathcal{X}_2) \in (x_2^C, \bar{x}] \cap \{x : \phi(x) > x\} \), and so on. If \( K \) were finite, the previous iteration would have to end after finitely many steps, say \( m \). But then \( \beta(\mathcal{X}_m) = \bar{x} \) and \( \beta(\mathcal{X}_m) \in \{x : \phi(x) > x\} \), giving \( \phi(\bar{x}) > \bar{x} \). The previous contradiction proves that \( K \) cannot be finite.

We proceed to show that \( K \) cannot be infinite either. The function \( U \) is continuous and, by (5), \( U(x_n^C) > U(x^*) \) for all \( x_n^C \in \mathcal{X}^C \cap (x^*, \bar{x}] \). Furthermore, as already pointed out above, \( \mathcal{X}^C \cap (x^*, \bar{x}] \) is a compact set. Therefore,

\[
\Delta := \min_{x_n^C \in \mathcal{X}^C \cap (x^*, \bar{x}]} U(x_n^C) - U(x^*) > 0. \tag{6}
\]

Next, \( U \) being continuous and \( \mathcal{X} \) compact, the function \( U \) is uniformly continuous on \( \mathcal{X} \). We can thus find \( \delta > 0 \) such that \( |U(x') - U(x)| < \Delta \) whenever \( |x' - x| < \delta \). By (6), we thus have

\[
U(x) > U(x^*), \text{ for all } x \text{ such that } |x - x_n^C| < \delta, x_n^C \in \mathcal{X}^C \cap (x^*, \bar{x}]. \tag{7}
\]

Now, since \((K, \beta)\) implements \( x^* \), we must have \( U(\beta(\mathcal{X}_i)) \leq U(x^*) \) for all \( \mathcal{X}_i \in K \). So (7) shows that each member of the sequence \( \mathcal{X}_1, \mathcal{X}_2, \ldots \) defined in the first part of the proof must have a length \( \delta \) or more. This in turn implies that said sequence can have no more than \( \frac{\bar{x} - x^*}{\delta} \) terms. Yet we showed previously that this sequence cannot be finite. This contradiction completes the proof of the theorem.

\[\square\]

B Appendix of Subsection 5.1

Proof of Theorem 2: We prove here the only if part of the theorem; the rest was proven in the text. Pick an arbitrary action \( x^* \) of the leader. Suppose that \( Q_{\leq}(x^*) \cap \{x : \phi(x) \leq x\} = \emptyset \). Applying Lemma 2 shows that any admissible pair \((K, \beta)\) in which the CST \( K \) contains only
intervals must be such that $\beta(\mathcal{X}_i) \in \{x : \phi(x) \leq x\}$ for every $\mathcal{X}_i \in K$ containing $\mathbf{x}$. This, in turn, implies that every I-plausible action belongs to $Q_>(x^*)$, whence $x^*$ cannot be I-plausible.

A similar argument shows that $Q_<(x^*) \cap \{x : \phi(x) \geq x\} = \emptyset$ implies that $x^*$ is not I-plausible.

Next, suppose that $Q_<(x^*) \cap \{x : \phi(x) \leq x\}$ and $Q_<(x^*) \cap \{x : \phi(x) \geq x\}$ are non-empty. Both $\phi$ and $U$ being continuous, the min and max of (3) are in this case well defined (since $\mathcal{X}$ is a compact set). Suppose that $\max Q_<(x^*) \cap \{x : \phi(x) \geq x\} < \min Q_<(x^*) \cap \{x : \phi(x) \leq x\}$, and pick

$$x^\dagger \in \{\max Q_<(x^*) \cap \{x : \phi(x) \geq x\}, \min Q_<(x^*) \cap \{x : \phi(x) \leq x\}\}.$$

(8)

Applying Lemma 2 shows that any admissible pair $(K, \beta)$ comprising an interval CST must be such that, for every $\mathcal{X}_i \in K$ containing $x^\dagger$, either (i) $\beta(\mathcal{X}_i) \in \{x \geq x^\dagger : \phi(x) \geq x\}$ or (ii) $\beta(\mathcal{X}_i) \in \{x \leq x^\dagger : \phi(x) \leq x\}$. So (8) gives $\beta(\mathcal{X}_i) \in Q_>(x^*)$. It ensues that $x^*$ cannot be I-plausible. □

**Proof of Proposition 1:** By Corollary 2, an action that belongs to the upper contour set of some Cournot action is I-plausible. Below we show that the converse is true too if $U$ is either quasi-convex or quasi-concave.

Suppose that $U$ is quasi-convex, and consider an action $x^*$ in the strict lower contour set of every Cournot action. Then $Q_<(x^*)$ is a convex set, and $\phi(x) \neq x$ for all $x \in Q_<(x^*)$. The intermediate value theorem shows that either $x < \phi(x)$ for all $x \in Q_<(x^*)$, or $x > \phi(x)$ for all $x \in Q_<(x^*)$. Either way, Theorem 2 shows that $x^*$ cannot be I-plausible.

Next, suppose that $U$ is quasi-concave, and consider an action $x^*$ in the strict lower contour set of every Cournot action. Then $Q_>(x^*)$ is a convex set, and $\phi(x) \neq x$ for all $x \in Q_<(x^*)$. This implies that, given $x \in Q_<(x^*)$, either (i) $\phi(x) > x$ and $x < x^C_n$ for all $x^C_n \in \mathcal{X}^C$, or (ii) $\phi(x) < x$ and $x > x^C_n$ for all $x^C_n \in \mathcal{X}^C$. We conclude using Theorem 2 that $x^*$ is not I-plausible. □
C Appendix of Subsection 5.2

Proof of Proposition 2: The if part of the proposition is evident. We prove here the only if part. Let \((x^*, y^*)\) be plausible, and \((K, \beta)\) be an admissible pair that implements \((x^*, y^*)\). Let \(B\) denote the image of the function \(\beta\), that is, \(B := \beta(K)\). For each action \(x \in X \setminus B\), let

\[
B_x := \bigcup_{\mathcal{X}_i \in K \text{ s.t. } x \in \mathcal{X}_i} \{\beta(\mathcal{X}_i)\};
\]

that is, \(B_x\) is the set of all leader’s actions played in some subgame containing \(x\). If \(K\) is not a partition, \(B_x\) may contain more than one element. For all \(x \in X \setminus B\), let \(\rho(x)\) be one element of \(B_x\). For all \(b \in B\), we can now define the set

\[
[b] := \{b\} \cup \{x \in X \mid \rho(x) = b\}.
\]

Finally, let

\[
K' := \{[x] \mid x \in B\},
\]

and \(\beta' : K' \to X\) such that \(\beta'([x]) = x\), for all \(x \in B\). By construction, \(K'\) is a partition of \(X\).

Since \((K, \beta)\) is admissible, we have \(\eta(\tilde{x}, x) \leq 0\) for all \(\tilde{x} \in B_x\), which implies that the pair \((K', \beta')\) is admissible as well. Moreover, it implements \((x^*, y^*)\) since \(\beta'(K') = B = \beta(K)\). ■

Lemma C.1. Suppose (RC1)–(RC3) hold. If \(u_2 u_{12} > 0\), then \(U\) is increasing over \([x, x^C]\). If \(u_2 u_{12} < 0\), then \(U\) is decreasing over \([x^C, \bar{x}]\).

Proof: We show the proof for the case in which \(u_2 > 0\) and \(u_{12} > 0\); the other cases are similar.\(^{34}\) Pick an arbitrary \(x < x^C\), and \(\varepsilon > 0\) sufficiently small that \(u(x + \varepsilon, R_F(x)) > u(x, R_F(x))\).\(^{35}\) Then, \(R_F\) being non-decreasing (since \(v_{12} > 0\) and \(u_2 > 0\):

\[
U(x + \varepsilon) = u(x + \varepsilon, R_F(x + \varepsilon)) \geq u(x + \varepsilon, R_F(x)) > u(x, R_F(x)) = U(x).
\]

\(^{33}\)Using the axiom of choice, we define a mapping \(\hat{\rho}\) that selects \(\hat{\rho}(B_x) \in B_x\) for each \(x \in X \setminus B\). Then, we define \(\rho : X \setminus B \to B\) by setting \(\rho(x) = \hat{\rho}(B_x)\).

\(^{34}\)Recall, \(u_2 > 0\) is shorthand notation for positive externalities and \(u_{12} > 0\) for supermodular payoffs.

\(^{35}\)The function \(u(\cdot, R_F(x))\) being strictly quasi-concave and maximized at \(\phi(x)\), it ensues that \(x < x^C\) implies \(\eta(x + \varepsilon, x) > 0\) for all sufficiently small \(\varepsilon > 0\).
Lemma C.2. Suppose (RC1)–(RC3) hold. Then
\[ S = \{ x : \eta(x^C, x) \leq 0 \} \cap \{ x : u(x^C, R_F(x)) \leq U(x^C) \}. \tag{9} \]

Proof: We show the proof of the lemma for the case \( u_2 > 0 \) and \( u_{12} > 0 \) (the other cases are similar). Recall that in this case \( S := \{ x : x \leq \gamma(x) \leq x^C \} \).

The function \( R_F \) being in this case non-decreasing (and, indeed, increasing in a neighborhood of \( x^C \) since \( y^C \in \text{int}(\mathcal{Y}) \)) and \( u_2 > 0 \), notice that \( u(x^C, R_F(x)) > u(x^C, R_F(x^C)) = U(x^C) \), for all \( x > x^C \).

So \( u(x^C, R_F(x)) \leq U(x^C) \) implies \( x \leq x^C \). Now consider \( x \leq x^C \) such that \( \eta(x^C, x) \leq 0 \). We will show that \( x \in S \). If \( x = x^C \) the previous claim is immediate, so pick \( x < x^C \). The function \( \eta(\cdot, x) \) is strictly quasi-concave, and maximized at \( \phi(x) > x \). As \( \eta(x, x) = 0 \geq \eta(x^C, x) \), we see by definition of \( \gamma(x) \) that \( x < \gamma(x) \leq x^C \). The right-hand side of (9) is thus contained in the set \( S \). The proof of the reverse inclusion is analogous. 

Lemma C.3. Suppose (RC1)–(RC3) hold, and \( S = \{ x^C \} \). Then all plausible actions belong to the upper contour set of \( x^C \) with respect to \( U \).

Proof: Reason by contradiction, and suppose that some action \( x^\ast \in Q_<(x^C) \) is plausible. Let \( (K, \beta) \) be an admissible pair that implements \( x^\ast \). Choose an element \( \mathcal{X}_i \) of the CST \( K \) such that \( x^C \in \mathcal{X}_i \). Using Lemma 1 yields \( \beta(\mathcal{X}_i) \in \{ x : \eta(x^C, x) \leq 0 \} \cap Q_<(x^\ast) \), and, since \( x^\ast \in Q_<(x^C) \),
\[ \beta(\mathcal{X}_i) \in \{ x : \eta(x^C, x) \leq 0 \} \cap Q_<(x^C). \tag{11} \]

\footnote{As \( \phi \) is continuous, notice that}
\[
\begin{align*}
\phi(x) > x & \quad \text{for } x < x^C, \\
\phi(x) < x & \quad \text{for } x > x^C.
\end{align*} \tag{10}
\]
In turn, (11) yields
\[ u(x^C, R_F(\beta(X_i))) \leq u(\beta(X_i), R_F(\beta(X_i))) = U(\beta(X_i)) < U(x^C). \] (12)

Coupling (11) and (12) gives
\[ \beta(X_i) \in \{ x : \eta(x^C, x) \leq 0 \} \cap \{ x : u(x^C, R_F(x)) < U(x^C) \}. \]

Applying Lemma C.2, we obtain \( \beta(X_i) \in S \setminus \{ x^C \} \), contradicting \( S = \{ x^C \} \).

**Lemma C.4.** Suppose (RC1)–(RC3) hold. Assume \( u_{12} > 0 \) and \( u_2 > 0 \). Consider an admissible pair \((K, \beta)\) which implements some action \( x^* \). Then, if \( x \in Q_>(x^*) \), we have \( \beta(X_i) < x \) for every \( X_i \in K \) which contains \( x \).

**Proof:** Let \( x \in Q_>(x^*) \), and pick an arbitrary \( X_i \in K \) containing \( x \). Reason by contradiction, and suppose that \( \beta(X_i) \geq x \). Then, \( R_F \) being non-decreasing (since \( v_{12} > 0 \)) and \( u_2 > 0 \), we obtain
\[ u(x, R_F(\beta(X_i))) \geq u(x, R_F(x)) > u(x^*, R_F(x^*)). \] (13)

Since \((K, \beta)\) is admissible, we also have
\[ u(\beta(X_i), R_F(\beta(X_i))) \geq u(x, R_F(\beta(X_i))). \] (14)

Coupling (13) and (14) yields
\[ u(\beta(X_i), R_F(\beta(X_i))) > u(x^*, R_F(x^*)). \]

The previous inequality contradicts the assumption that \((K, \beta)\) implements \( x^* \).

**Lemma C.5.** Suppose (RC1) holds. Let \((K, \beta)\) be an admissible pair. If \( \beta(X_i) < \min\{x^C, x\} \) for some \( X_i \in K \) which contains \( x \), then \( \gamma(\beta(X_i)) \in (\beta(X_i), x] \).

**Proof:** Pick \( x \in X \), and \( X_i \in K \) containing \( x \). Since \((K, \beta)\) is admissible:
\[ \eta(x, \beta(X_i)) \leq 0. \] (15)

Now suppose that \( \beta(X_i) < \min\{x^C, x\} \). In this case, the strictly concave function \( \eta(\cdot, \beta(X_i)) \) attains (by virtue of (10)) a maximum at \( \phi(\beta(X_i)) > \beta(X_i) \). From (15) and the fact that \( \beta(X_i) < x \) we obtain (by definition of \( \gamma \)) \( \beta(X_i) < \gamma(\beta(X_i)) \leq x. \)
Proof of Theorem 3: Start with the case $S = \{x^C\}$. Combining Corollary 1, Proposition 2, and Lemma C.3 shows that the set of simply-plausible actions, the set of I-plausible actions, the set of P-plausible actions, and the set of plausible actions all coincide with the upper contour set of $x^C$.

The remainder of the proof deals with the case $S \supseteq \{x^C\}$. Below, assume $u_{12} > 0$ and $u_2 > 0$ (the other cases are analogous). Recall that in this case $S := \{x : x \leq \gamma(x) \leq x^C\}$. The function $\gamma$ being continuous, $S$ is a compact set. By Lemma C.1, we can thus find $\hat{x} \in S$ with $\hat{x} < x^C$ and

$$U(\gamma(\hat{x})) = \min_{x \in S} U(\gamma(x)).$$

(16)

To shorten notation, let $\hat{\gamma} := \gamma(\hat{x})$; as $\hat{x} < x^C$, note that, by definition of $\gamma$,

$$\hat{x} < \hat{\gamma} \leq x^C.$$

(17)

We proceed to show that (a) all actions in $Q_{\geq}(\hat{\gamma})$ are P-plausible, and (b) any plausible action belongs to $Q_{\geq}(\hat{\gamma})$.

All actions in $Q_{\geq}(\hat{\gamma})$ are P-plausible. We know by Corollary 1 that all actions in $Q_{\geq}(x^C)$ are simply plausible. So pick an action $x^* \in Q_{\geq}(\hat{\gamma}) \setminus Q_{\geq}(x^C)$ (if there exists none, we are done). Define

$$\mathcal{X}_1 := \{\hat{x}\} \cup Q_{>}(x^*),$$

and let $K$ denote the partition of $\mathcal{X}$ made up of $\mathcal{X}_1$, and only singletons besides $\mathcal{X}_1$. Lastly, let $\beta : K \to \mathcal{X}$ be given by $\beta(\mathcal{X}_1) = \hat{x}$ and $\beta(\{x\}) = x$ for all $x \in \mathcal{X} \setminus \mathcal{X}_1$. We now show that $(K, \beta)$ constitutes an admissible pair; notice that this amounts to showing that

$$\eta(\bar{x}, \hat{x}) \leq 0, \quad \text{for all } \bar{x} \in \mathcal{X}_1.$$  

(18)

As $x^* \in Q_{\geq}(\hat{\gamma})$, any $\bar{x} \in Q_{>}(x^*)$ belongs to $Q_{\geq}(\hat{\gamma})$. On the other hand, since $\hat{\gamma} \leq x^C$ (see (17)), Lemma C.1 shows that every $\bar{x} \in Q_{>}(x^*)$ satisfies $\bar{x} \geq \hat{\gamma}$. Now, the function $\eta(\cdot, \hat{x})$ is strictly quasi-concave, with $\eta(\hat{x}, \hat{x}) = \eta(\hat{\gamma}, \hat{x}) = 0$; it thus follows from (17) that $\eta(\bar{x}, \hat{x}) \leq 0$ for all $\bar{x} \geq \hat{\gamma}$. Combining the previous observations establishes (18); so $(K, \beta)$ is admissible.
Finally, coupling (17) and Lemma C.1 yields $U(\hat{\gamma}) > U(\hat{\delta})$, giving in turn $U(x^*) > U(\hat{\delta}) = U(\beta(\mathcal{X}_1))$ (since $x^* \in Q_{\geq}(\hat{\gamma})$). We conclude that $(K, \beta)$ implements $x^*$, since $\mathcal{X} \setminus \mathcal{X}_1 \subset Q_{\leq}(x^*)$. *All plausible actions belong to $Q_{\geq}(\hat{\gamma})$. Reason by contradiction, and suppose that some plausible action $x^*$ belongs to $Q_{<}(\hat{\gamma})$. Combining (17), Lemma C.1, and the fact that $U$ is continuous shows that we can find an action, say $x^\dagger$, such that:

\[ x^\dagger < \hat{\gamma}, \]  

(19)

and

\[ x^\dagger \in Q_{<}(x^*) \cap Q_{<}(\hat{\gamma}). \]  

(20)

Now consider a pair $(K, \beta)$ which implements $x^*$, and $\mathcal{X}_i$ an element of the CST $K$ containing $x^\dagger$. By virtue of (20), applying Lemma C.4 shows that

\[ \beta(\mathcal{X}_i) < x^\dagger. \]  

(21)

On the other hand, (17) and (19) show that

\[ x^\dagger < \hat{\gamma} \leq x^C. \]

Hence, Lemma C.5 gives

\[ \beta(\mathcal{X}_i) < \gamma(\beta(\mathcal{X}_i)) \leq x^\dagger < \hat{\gamma} \leq x^C. \]  

(22)

We thus obtain, firstly,

\[ \beta(\mathcal{X}_i) \in S, \]  

(23)

and, secondly (using Lemma C.1),

\[ U\left(\gamma(\beta(\mathcal{X}_i))\right) < U(\hat{\gamma}). \]  

(24)

The combination of (23) and (24) contradicts (16). Therefore, every plausible action must belong to $Q_{\geq}(\hat{\gamma})$. ■

**Proof of Proposition 3:** By definition of $\gamma$: $\eta(\gamma(x), x) = 0$ for all $x$ in some neighborhood $O$ of $x^C$. We thus have

\[ u(\gamma(x), R_F(x)) = u(x, R_F(x)), \quad \forall x \in O. \]  

36
Differentiating the previous expression with respect to \( x \) yields

\[ u_1(\gamma(x), R_F(x))\gamma'(x) + u_2(\gamma(x), R_F(x)) R'_F(x) = u_1(x, R_F(x)) + u_2(x, R_F(x)) R'_F(x), \]

and, therefore,

\[ \gamma'(x) = \frac{u_1(x, R_F(x)) + R'_F(x)[u_2(x, R_F(x)) - u_2(\gamma(x), R_F(x))]}{u_1(\gamma(x), R_F(x))}, \quad \forall x \in O \setminus \{x^C\}. \tag{25} \]

The numerator and denominator on the right-hand side of (25) tend to 0 as \( x \to x^C \). Then, by virtue of L'Hôpital’s rule and using the fact that \( \gamma(x) \to x^C \) as \( x \to x^C \):

\[ \lim_{x \to x^C} \gamma'(x) = \lim_{x \to x^C} \frac{u_1(x, R_F(x)) + 2u_12(x, R_F(x)) R'_F(x) - u_12(x, R_F(x)) R'_F(x) \gamma'(x)}{u_1(\gamma(x), R_F(x)) \gamma'(x) + u_12(\gamma(x), R_F(x)) R'_F(x)}. \tag{26} \]

On the other hand, in a neighborhood of \( y = y^C \):

\[ R'_L(y) = \frac{-u_12(R_L(y), y)}{u_11(R_L(y), y)}. \]

Therefore,

\[ R'_L(y^C) = \frac{-u_12(x^C, y^C)}{u_11(x^C, y^C)} = \lim_{x \to x^C} \frac{-u_12(x, R_F(x))}{u_11(x, R_F(x))} = \lim_{x \to x^C} \frac{-u_12(\gamma(x), R_F(x))}{u_11(\gamma(x), R_F(x))}. \tag{27} \]

Combining (27) with (26) gives

\[ \gamma'(x^C) = \frac{1 - 2R'_L(y^C) R'_F(x^C) + R'_L(y^C) R'_F(x^C) \gamma'(x^C)}{\gamma'(x^C) - R'_L(y^C) R'_F(x^C)}. \]

So \( \gamma'(x^C) \) is a solution of

\[ Z(Z - 2\alpha) = 1 - 2\alpha, \]

where \( \alpha := R'_L(y^C) R'_F(x^C) \). So either \( \gamma'(x^C) = 1 \) or \( \gamma'(x^C) = 2\alpha - 1 \), whence \( \gamma'(x^C) > 0 \) if \( R'_L(y^C) R'_F(x^C) > 1/2 \).

Now suppose that \( u_12u_2 > 0 \) (the other case is similar), so that \( S = \{ x : x \leq \gamma(x) \leq x^C \} \).

If \( R'_L(y^C) R'_F(x^C) > 1/2 \), then \( \gamma'(x^C) > 0 \). This in turn implies the existence of \( x < x^C \) such that \( x < \gamma(x) < x^C \). Such an \( x \) belongs to \( S \), so Lemma C.1 enables us to conclude that \( U < U(x^C) \). \hfill \blacksquare