GAMES AND ECONOMIC BEHAVIOR **10,** 255–283 (1995)

Ratifiable Mechanisms: Learning from Disagreement

PETER C. CRAMTON

University of Maryland, College Park, Maryland 20742

AND

THOMAS R. PALFREY

California Institute of Technology, Pasadena, California 91125

Received March 26, 1991

In a mechanism design problem, participation constraints require that all types prefer the proposed mechanism to some status quo. If equilibrium play in the status quo mechanism depends on the players' beliefs, then the inference drawn if someone objects to the proposed mechanism may alter the participation constraints. We investigate this issue by modeling the mechanism design problem as a two-stage process, consisting of a ratification stage followed by the actual play of the chosen game. We develop and illustrate a new concept, *ratifiability*, that takes account of inferences from a veto in a consistent way. *Journal of Economic Literature* Classification Numbers: C700, C780, D820.

1. INTRODUCTION

Mechanism design is a powerful theory for studying incentive problems in settings where privately informed decision makers have conflicting interests. In a typical application of the mechanism design approach, the analyst is able to characterize the set of outcomes that are attainable by the agents, recognizing each agent's voluntary participation and incentive to misrepresent information that is privately known. In addition, it is often possible to characterize "optimal" mechanisms—mechanisms that are efficient in either an ex ante or interim sense.

Despite these significant virtues, one can argue that the mechanism design approach identifies too large a set of attainable outcomes, because it allows the agents to make unreasonable commitments, both before and after the selection of a mechanism. Commitment becomes an issue if information is leaked during the selection or implementation of a mechanism. The information leakage alters the incentive problem faced by the agents, creating opportunities for renegotiation. An inability to commit to not using leaked information typically reduces the set of attainable outcomes, because additional incentive constraints must be satisfied. Several authors have focused on the commitment problem faced by agents as a result of information leakage during the implementation of a mechanism. See, for example, Ausubel and Deneckere (1989), Caillaud and Hermalin (1989), Cramton (1985), and Green and Laffont (1985, 1987).

Others have addressed the related problem of information leaked during the process of selecting a mechanism. For example, Myerson (1983) and Maskin and Tirole (1990, 1992) analyze the problem that a privately informed principal has in selecting a mechanism when recognizing that the selection may reveal information to the subordinates. Ideally, the principal would like to condition her choice of mechanism on her private information, but to do so would reveal information and hence make the chosen mechanism invalid, assuming it is based on the prior beliefs of the subordinates.

One way around this problem of information leakage during the selection of a mechanism is to assume that the agents select the mechanism in the ex ante stage, before they have their private information. But even if selection is done by uninformed agents, once the agents learn their private information—the interim stage—they may have an incentive to renegotiate to a different mechanism, as Holmström and Myerson (1983) demonstrate. Again, if the agents are unable to commit to not renegotiating, then the set of attainable outcomes is further constrained.

Here we focus on each agent's decision to participate in the mechanism and explore the possibility that a refusal to participate may reveal information. To isolate this information leakage problem from the others, we assume the following.

- The proposed mechanism is selected by an uninformed third party, so no information is revealed by its selection.

- The agents can commit to not renegotiating to an alternative proposal at the interim stage, thus avoiding the durability problem discussed in Holmström and Myerson (1983) and Crawford (1985).

0899-8256/95 $12

- The agents voluntarily decide to participate in the mechanism at the interim stage. If participation is unanimous, then the agents are bound by the mechanism; otherwise, they play a *status quo mechanism*, possibly with altered beliefs. In either case, the agents are ultimately committed to the mechanism or the status quo and cannot renegotiate during implementation.

It is important that an uninformed third party selects the mechanism, rather than the agents in the exante stage. Otherwise, the agents would be able to commit to the mechanism in the ex ante stage, so only ex ante individual rationality, rather than interim individual rationality, would be required. The commitment assumption, while strong, is not unreasonable in an environment where contracts can be enforced.[1]

Our innovation over the standard mechanism design approach is to allow the possibility of learning from the agents' participation decisions. This is accomplished by modeling the ratification process as a two-stage game: a ratification stage in which every player has an opportunity to veto the proposed mechanism, followed by an implementation stage in which either the proposed mechanism or the status quo is adopted, depending on whether the proposed mechanism was unanimously ratified in the first stage.[2] A mechanism is *ratifiable* relative to the status quo if unanimous ratification is a sequential equilibrium of the two-stage game, where beliefs following disagreement are required to satisfy consistency conditions similar to those proposed by Farrell (1985) and Grossman and Perry (1986).

In many mechanism design settings learning from disagreement is not an issue, since the status quo outcome is not affected by a change in the players' beliefs. For example, in bilateral trade a natural status quo would be no trade (Myerson and Satterthwaite, 1983). But in many other situations (and perhaps even in the bilateral trade case) equilibrium play in the status quo mechanism may depend on the players' revised beliefs, *conditioned on the failure to ratify the proposed mechanism.* If this is the case, then ratification gives a player an opportunity to signal information, which in turn may affect the desirability of playing the status quo mechanism. Hence, an agent s participation decision depends on what he expects others to infer from his veto of the proposed mechanism.

Given a definition of ratifiability that specifies what alternative mechanisms can be unanimously approved against a status quo mechanism when beliefs are not passive, we go on to explore the question of *secure* status quo mechanisms, mechanisms which are immune to the ratification of any alternative mechanism. That is, a mechanism (or an equilibrium allocation of the mechanism) is secure if there does not exist an alternative mechanism that could be unanimously ratified against it, given reasonable updating in the event of a veto. Clearly the concept of security/ratifiability is related to Holmström and Myerson's (1983) concept of durability/rejectability and Legros' (1991) stable allocations. However, because they focus on learning that may arise as a result of successful ratification and we focus on learning that results from unsuccessful ratification, the concepts diverge. Our approach also differs because we consider nonanonymous ratification procedures while these other approaches are limited to anonymous voting using a secret ballot. These differences are discussed later in the paper, where we also establish the existence of interim efficient and secure mechanisms in a class of environments.

We now give three examples of mechanism design problems where learning during the ratification process is relevant.

Cartels (Cramton and Palfrey, 1990). The firms in an industry wish to form a cartel. Each firm has private information about its marginal cost, as in Roberts (1985). Under the proposed (monopoly) mechanism, the firms announce their costs, the firm with the lowest cost produces the monopoly level of output, the others produce nothing, and the monopoly revenue is split among the firms based on the vector of reported costs so as to elicit truthful reports. If the proposed mechanism is not unanimously ratified, a status quo mechanism consisting of (Bayesian) Cournot competition ensues: each firm simultaneously selects output to maximize profit given its own costs and its belief about the costs of the other firms and how much they will produce. But what beliefs should the firms hold in the event of disagreement? These beliefs are important, because they determine a firm's payoff in the

[1] Renegotation becomes a problem if the parties end up in a situation where everyone agrees to change the mechanism. This state, however, can apparently be avoided if communication is limited or if the agents include in their initial contract a clause requiring all parties to pay large sums of money to a third party in the event the contract is renegotiated. A court interested in efficient contracting should enforce such clauses. The argument is not so simple, however, since the third party may agree to tear up the contract for a small fee, since if the contract is not torn up the parties will never renegotiate and so the third party will get nothing. The argument, therefore, rests on the assumption that the third party is not susceptible to such bribes.

[2] Maskin and Tirole's (1990, 1992) analysis of the informed principal problem also has a ratification stage, but our models are quite different. In Maskin and Tirole an uninformed agent chooses to ratify the mechanism. Information is revealed from the informed principal's selection of a mechanism, rather than the ratification process. In our model, an uninformed principal selects the mechanism and then informed agents choose to ratify. Information is revealed from the ratification process, but not from the choice of mechanism.

status quo, and hence whether it should ratify the proposed mechanism. Individual rationality critically depends on what is inferred from a (possibly unexpected) veto of the proposed monopoly mechanism.

One possibility, consistent with the standard mechanism design approach, is that nothing is learned from disagreement. With two firms, if the proposed mechanism has been designed to be incentive compatible and interim individually rational (assuming nothing is learned), then all types of firms have an interest in ratifying the mechanism, so disagreement is a zero-probability event. Passive beliefs (no learning) will support the proposed mechanism as a sequential equilibrium in this ratification game.

We argue, however, that passive updating is implausible. The ratifying firm, surprised by the disagreement of the other, should try to rationalize the deviation by identifying a set of types (called the *veto set*) that could stand to benefit from vetoing. The veto set is credible if those in the veto set prefer the status quo mechanism to the alternative mechanism, and those not in the veto set prefer the mechanism to the status quo, where the status quo payoffs are calculated at an equilibrium with the updated belief that the vetoer is in the veto set. In this way, we restrict beliefs about a vetoing firm to be a credible veto set of types (if one exists).[3]

In our cartel example, it is low-cost firms that benefit the least from participating in the cartel. Might a lost-cost firm stand to gain by vetoing if by doing so it signalled that it had a low cost? If the others believe that the vetoing firm has a low cost, then they expect this firm to produce a lot and so they optimally respond by reducing their output, thus increasing the status quo payoff to the vetoing firm. Since low-cost firms gain the least from participation, this increase in the status quo payoff may be sufficient for them to prefer the status quo to the mechanism, but high-cost firms (who gain a great deal from the mechanism) would still prefer the mechanism to the status quo even if by vetoing a high-cost firm is able to convince the others that it is a low-cost firm. Here learning from disagreement has the effect of strengthening the participation constraints by improving the status quo payoffs for the vetoer, and therefore reduces the set of attainable outcomes. In particular, the monopoly mechanism is not ratifiable. The degree of collusion is limited by a low-cost firm's credible threat of overproduction.

Litigation (Spier, 1988; Spulber, 1989). A plaintiff and defendant are engaged in pretrial negotiation. The plaintiff has private information about her level of damages and the defendant has private information about his liability. They would like to settle their dispute without a trial because a trial is costly to both. The proposed mechanism specifies a settlement amount and a probability of going to court as a function of the reports of private information. The status quo is going to court if the plaintiff believes bringing the case to court is profitable. If the case does go to court, each player's private information is revealed during the costly discovery process and then the court awards to the plaintiff her damages times the degree of liability of the defendant.

Note that the defendant's status quo payoff depends on the plaintiff's decision to take the case to court, which in turn depends on the plaintiff's belief about the liability of the defendant. If the plaintiff's damages are sufficiently small or she believes the defendant is not sufficiently liable, then the plaintiff prefers not to take the case to court, since the cost of discovery and trial is greater than the expected award.

What type of defendant gains the least from participating in the mechanism? If the proposed mechanism is incentive compatible and does not suggest going to court for plaintiffs with damages that are too small to make going to court profitable, then the least liable type of defendant gains the least from participating in the mechanism relative to the status quo. Vetoing the mechanism may be a credible signal of low liability, since signaling low liability reduces the probability that the plaintiff will take the case to court and hence raises the defendant's status quo payoff. So long as defendants with high liability still gain more from the mechanism than the improved status quo, the low liability types form a credible veto set. Learning from disagreement in this example would have the same effect as in the cartel example: the learning improves the status quo payoffs for the vetoer, and so the set of attainable outcomes is reduced.

Arbitration. A buyer and a seller are negotiating a price for an object owned by the seller. The buyer has private information about his reservation price, as does the seller. A mechanism specifies a trading price and a probability of trade as a function of the private information of the two parties. The status quo is either no trade or arbitration if both agree to arbitrate. Arbitration is a costly verification procedure similar to the courts in the litigation example. The arbitrator, through costly investigation, verifies the traders' reservation prices and then has them split any gains from trade equally. Incentive compatibility typically implies that the highest valuation seller and the lowest valuation buyer gain the least from participating in such a mechanism. Can vetoing be a credible signal of a high value for the seller? In this case, signaling a high type is bad for the seller in the status quo mechanism, since it reduces the probability that the buyer will be willing to bear the cost of arbitration. A low-value seller would have no interest in

[3] Other restrictions on beliefs following a veto are considered as well. These restrictions correspond to alternative equilibrium refinements.

signaling a high type by vetoing, since the low type benefits more from the mechanism than a high type. Hence, unlike the last two examples, here learning from disagreement has the effect of reducing the status quo payoffs for the vetoer, and thereby expanding the set of individually rational mechanisms.

There are numerous other environments where the learning-from-disagreement problem arises, such as collusion in auctions, regulation, and multi-agent incentive schemes, the adoption of voting rules in legislatures and committees, and elsewhere. The central feature of these applications is that the outcome of the status quo mechanism depends on actions that the agents take based on their beliefs about private information. Thus, what is learned from the veto of a proposed mechanism can affect the vetoer's status quo payoff and therefore the agent's decision to ratify. Individual rationality, to the extent that it embodies the play of a default mechanism following unsuccessful ratification, depends on what is learned from disagreement. To answer the question, "What are the relevant individual rationality constraints?", we must first answer the question, "What beliefs should the agents have following disagreement?"

Section 2 presents the general model and definitions. Section 3 applies the general model to the cartel problem studied in Cramton and Palfrey (1990). The cartel example serves to illustrate that even ex post efficient mechanisms that are feasible, incentive compatible, and individually rational without learning may be eliminated by refinements that place reasonable restrictions on beliefs following disagreement. Section 4 compares the implications of our model of learning from disagreement with several alternatives from the literature on equilibrium refinements. Section 5 defines secure mechanisms, compares the notions of security and durability, and proves existence of interim efficient and secure mechanisms for a class of environments. Proofs not presented in the main text appear in the Appendix.

2. RATIFIABLE MECHANISMS

Consider a mechanism design problem with n individuals, indexed by $i \in N = \{1,\ldots, n\}$, that must make a decision $d \in D$. Each individual i has private information $t_i \in T_i$, representing a realization of all of i's information that is not common knowledge. For simplicity, we assume that the sets D and $T = T_1 \times \cdots \times T_n$ are finite. In addition, we assume types are independent: t_i is drawn from the distribution p_i, independently of t_j for $j \neq i$ and this is common knowledge. Player i's ex post utility u_i: $D \times T \to \Re$ depends on the decision and vector of types $t=(t_1,\ldots,t_n) \in T$.[4] A decision rule d: $T \to D$ maps each vector of types into a decision. The decision rule d is implemented as a direct mechanism: the players simultaneously report their types t', possibly dishonestly, and then the decision $d(t')$ is adopted. An uninformed third party, the mechanism designer, must propose a particular decision rule, recognizing incentive and participation constraints. Incentive constraints are dealt with by requiring that the decision rule d be incentive compatible. $d(t)$ is incentive compatible if, in the direct mechanism, honesty is a best response for each player given that the others are honest. By the revelation principle, there is no loss of generality in requiring that the designer propose an incentive compatible decision rule. In what follows, d will always refer to an incentive compatible decision rule. Participation constraints are handled by requiring that every type of every player prefers to participate in the mechanism d than to play the *status quo mechanism* G, with finite message space $\mathtt{M}$ and outcome function g.

In the standard mechanism design approach, the participation constraints are quite simple, because the payoffs in the status quo mechanism, the alternative to participation, do not depend on strategic actions. Often the status quo simply specifies a constant value, say U^0. In that case, interim participation constraints are that the interim utility from the mechanism for each type of each player be at least U^0.

Here we allow for a status quo that depends on strategic actions. Participation constraints are considered by analyzing a two-stage ratification game. In the first stage, the players each vote simultaneously either for or against the proposed mechanism d. A strategy for i in the first stage specifies the probability each type t_i of i votes against d, denoted $u_i(t_i)$. An outcome to the first stage indicates the players $M \subseteq N$ that vetoed d. In the second stage, d is implemented if it is unanimously ratified ($M = \varnothing$), otherwise the status quo mechanism G is played with the knowledge that players M vetoed d.[5] Player i, in deciding whether to veto the mechanism d, considers how the others' beliefs about i might change as a result of the veto. This change in belief may alter i's payoff in the status quo

[4] In parts of the paper (for example, parts of Sections 3-5) we place restrictions on the utility functions of the players, $u_i(d, t)$. Elsewhere, in the absence of such restrictions, the assumption of independence could be dispensed with, since dependent-type models can be transformed into independent-type models by an appropriate transformation of the utility functions (Myerson, 1985).

[5] We assume the vote is public, since we are especially concerned with situations where one type of player has an incentive to publicly announce displeasure with δ, because of the information the announcement conveys.

mechanism, since the equilibrium played in may depend on the information revealed by the veto. What we attempt to do here is to link together 's decision to ratify with rational expectations about the outcome in the continuation game defined by

Because ratification requires a unanimous vote, we call *d*

of the two stage game (vote in stage 1; play either *G* *d* in stage 2) in which no type of any player vetoes. *i*, we must specify the beliefs of all the other players if unilaterally deviates from this ratification equilibrium by vetoing. We denote this belief about 's type following a veto by *i* m_i [6]

In particular, suppose that if alone vetoes *d* *i*'s type $_i$ is given by the distribution $_i$, *i* update their beliefs

i's type to be $_i$. Let (m) denote an equilibrium of *G* $U_i^G(t_i, s(m_i))$

to be the interim payoff to t_i in the equilibrium $s(m_i)$ of *G* when only *i* vetoes and the others infer *i*'s type is given by m_i. Finally, let $U_i^d(t_i)$ be t_i's interim payoff in the mechanism *d* (which corresponds to the equilibrium path for the equilibrium in which all types of all players vote for ratification) and define $U_i(t_i, s(m_i)) = U_i^d(t_i) - U_i^G(t_i, s(m_i))$ to be t_i's net benefit from *d* relative to *G*, if a veto by *i* results in the belief m_i and $s(m_i)$ is played in the continuation game. Let $m = \{m_1,\ldots,m_n\}$ be a collection of veto beliefs following any veto by a single player when all other players vote for ratification, and let $\sum(m) = \{s(m_1),\ldots,s(m_n)\}$ denote a corresponding collection of equilibria in the resulting play of the status quo mechanism with those beliefs.

DEFINITION. *d* is individually rational relative to the status quo *G* if there exists *m* and $\Sigma(m)$ such that for each *i* and each t_i, $U_i(t_i, s(m_i)) \geq 0$.

Our definition is the standard definition of individual rationality, but now the payoff in the status quo can depend on the inferences other players make from *i*'s unilateral veto of the mechanism. This is a weak definition of individual rationality, since it does not impose any restriction on the belief *m*i following a veto by *i*. Our interest in this definition stems from the following fact.

PROPOSITION 1. *d is incentive compatible and individually rational relative to G if and only if unanimous ratification of d followed by truthful revelation in the direct mechanism d is a sequential equilibrium in the ratification game.*

Proof (Only if). If in equilibrium all types of all players ratify *d* then Bayes' rule implies that the prior beliefs are unchanged when the direct mechanism constructed to implement *d* is played. Since *d* is incentive compatible, truthful revelation in this game is a best response following unanimous ratification. Consider any player *i* and $j \neq i$. Let *j's* belief if *i* deviates by vetoing *d* be m_i, and let all players follow $s(m_i)$ in the continuation game. Given this, a veto by *i* in the ratification stage is unprofitable for any type t_i, since the fact that *d* is individually rational relative to *G* implies that $U_i(t_i, s(m_i)) \geq 0$.

(If) *d* is incentive compatible, since, following unanimous ratification, truthful revelation in the direct mechanism is part of a sequential equilibrium (and so must be a best response). Let m_i denote the off-the-equilibrium-path belief about *i*'s type if *i* vetoes *d.* (Recall that consistency of beliefs implies that this belief must be the same for all players other than *i*.) Let $s(m_i)$ denote the equilibrium following a veto by *i*. Since unanimous ratification is a best response, t_i's net gain from ratifying must be nonnegative: $U_i(t_i, s(m_i)) \geq 0$. Hence, *d* is individually rational relative to *G*.

2.1. *Credible Veto Beliefs and Ratifiability*

We now explore restrictions on beliefs in the ratification game, based on a refinement proposed by Grossman and Perry (1986). In particular, we suppose that if *i* vetoes the mechanism, then the others will try to rationalize the veto by identifying a veto belief that is consistent with *i*'s incentive to veto. The veto belief m_i is induced from the prior p_i and the veto probabilities u_i. Define the veto set $V_i \subseteq T_i$ as those types that veto with positive probability: $V_i = \{t_i \in T_i \mid u_i(t_i) > 0\}$. We require that V_i is nonempty.

[6] Because all players other than *i* share a common prior about *i*'s type, the consistency of beliefs requirement of sequential equilibrium implies that all players $j \neq i$ must have the same belief following any action by *i* in the voting stage.

DEFINITION. A probability distribution m_i on T_i is a *credible veto belief* about i relative to the decision rule d and status quo mechanism G if there exists a continuation equilibrium $s(m_i)$ and veto probabilities $u_i(\cdot)$ such that m_i, $s(m_i)$, and $u_i(\cdot)$ together satisfy:

(i) $u_i(t_i) > 0$ for some $t_i \in T_i$,
(ii) $u_i(t_i) = 1$ for all $t_i \in T_i$ such that $U_i(t_i, s(m_i)) < 0$,
(iii) $u_i(t_i) = 0$ for all $t_i \in T_i$ such that $U_i(t_i, s(mi)) > 0$, and
(iv) m_i satisfies Bayes' rule given p_i and u_i:

$$m_i(t_i) = \begin{cases} \dfrac{p_i(t_i)u_i(t_i)}{\sum_{t_i \in V_i} p_i(t_i)u_i(t_i)} & \text{for } t_i \in V_i \\ 0 \text{ for } t_i \notin V_i\ . \end{cases}$$

The set of types V_i that veto with positive probability in m_i is called a *credible veto set.*

The definition requires that (ii) those types that benefit from a veto are assumed to veto the mechanism, and (iii) those types that lose from a veto are assumed not to veto the mechanism. There is no restriction on u_i for types that are indifferent between vetoing and not.

If m_i is a credible veto belief, then the others can rationalize i's veto by believing i's type is distributed according to m_i. If for every credible veto belief there is some type t_i that strictly gains by vetoing, then d is not ratifiable relative to G. A *credible veto belief profile* is a vector of beliefs $m = (m_1,\ldots, m_n)$ such that m_i is a credible veto belief for each i.

DEFINITION. An incentive compatible decision rule d is *ratifiable against* G if d is individually rational relative to G and for all i either

(i) there does not exist a credible veto belief for i, or

(ii) there exists a credible veto belief m_i and a corresponding equilibrium $s(m_i)$ of G under beliefs m_i such that $U_i(t_i, s(m_i)) = 0$ for all $t_i \in V_i$.

If after a veto the players' beliefs are restricted to credible veto beliefs when one exists, then we should require that the mechanism be ratifiable. For any mechanism that is not ratifiable, regardless of the credible veto belief, there is some type of player that strictly benefits from vetoing.

Remark 1. If d is ratifiable against G, then d is individually rational relative to G, by definition. Hence, from Proposition 1, unanimous ratification of d followed by truthful revelation in the direct mechanism d is a sequential equilibrium in the ratification game. An individually rational mechanism, however, need not be ratifiable, as we show by example later in the paper. Ratifiability is a stronger requirement than individual rationality.

Remark 2. It is certainly possible that d and G are such that no credible veto beliefs exist and the mechanism is individually rational relative to G. In this case, d is ratifiable, the participation constraints are not binding, and the beliefs following disagreement are indeterminate (and inconsequential). However, in many settings, the participation constraints are binding, provided the incentive problem is severe enough and the mechanism designer chooses d to be optimal in some sense. Such a mechanism typically is associated with the existence of a credible veto belief for each i with the property that all types in V_i are indifferent between ratifying and vetoing. In this case, the beliefs following disagreement are required to put all the weight on those types that gain the least from participating in the mechanism d relative to G.

Remark 3. A well-known difficulty with the perfect sequential equilibrium (PSE) concept is that PSE sometimes fail to exist. Because of the similarity of our notion of ratifiability to PSE, one might worry that, for some G, the set of ratifiable mechanisms relative to G may be empty. This, however, is never the case. Let $s(p_i)$ be an equilibrium to G under the prior beliefs. Let d be the decision rule generated by $s(p_i)$. Then d is always ratifiable against G, since for each i, p_i is a credible veto belief with $U_i(t_i, s(p_i)) = 0$ for all $t_i \in T_i$. This existence result stems from our ability to pick d. It is therefore quite different from the existence of a PSE for a fixed game.

Remark 4. It is possible that there is more than one way to rationalize a veto by player i.[7] In that case, there will be more than one credible veto belief for i. In fact, it is possible that for one of i's credible veto beliefs, say m_i, $U_i(t_i, s(m_i)) = 0$ for all $t_i \in V_i'$, but for another of i's credible beliefs, say $m_i', U_i(t_i, s(m_i')) < 0$ for some $t_i \in V_i'$. In this case, the mechanism would be ratifiable, as long as appropriate credible veto beliefs could be found for all $j \neq i$, even though there existed a credible veto belief for i, m_i', that would make some i-types better off under the status quo than under the proposed mechanism. This issue of multiple credible veto beliefs will be addressed next. The question is whether part (ii) of the definition of ratifiability should be required to hold for all credible veto beliefs.

2.2. *Strong Ratifiability*

As pointed out in Grossman and Perry (1986), there is a subtle difference between their definition of PSE and a related equilibrium refinement proposed by Farrell (1985). Farrell's refinement, neologism proof, differs from PSE because it requires an equilibrium to be supported by all, rather than one, credible updating rules for rationalizing observations off the equilibrium path. If neologism proofness is applied in our framework instead of PSE we get a strengthening of our definition of ratifiability.

DEFINITION. An incentive compatible mechanism d is *strongly ratifiable* against G if d is individually rational relative to G and for all i either

(i) there does not exist a credible veto belief, or

(ii) for every credible veto belief m_i, $U_i(t_i, s(m_i)) = 0$ for all $t_i \in V_i$.

In other words, if d is strongly ratifiable against G, then there does not exist a credible veto belief for any player such that some type in that veto belief strictly prefers the status quo to the mechanism. Any mechanism that is strongly ratifiable is ratifiable because part (ii) of the definition must hold for all credible veto beliefs, rather than just one.[8]

The two definitions of ratifiability coincide if there is at most one credible veto belief for any i. This will be the case in the cartel example studied in Section 3. Strong ratifiability, however, will be useful in Section 5, when we consider status quo mechanisms that cannot be overturned by a strongly ratifiable alternative.

In contrast to the definition of ratifiable, it may be that the status quo G is not strongly ratifiable against itself (or, more precisely, G is not strongly ratifiable against a decision rule produced at some equilibrium of G). There may be a credible veto belief m_i such that at least one type t_i strictly prefers an equilibrium in the status quo mechanism with the revised beliefs m_i relative to the status quo with beliefs p_i. Such a mechanism would not be impervious to allowing players to make binary preplay announcements ("Veto" or "Ratify") that may communicate information about their types. Thus, we can interpret the idea of a mechanism being strongly ratifiable against itself as permitting a special sort of preplay communication before G is carried out.[9] In this way we see that strong ratifiability implies some degree of cheap-talk proofness.

3. RATIFIABILITY OF CARTEL AGREEMENTS WITH A COURNOT THREAT

We now turn to an example from Cramton and Palfrey (1990) that illustrates how learning from disagreement can affect the set of ratifiable mechanisms. Whether learning from disagreement strengthens or weakens the participation constraints depends in general on the particular mechanism d. In our example, even though the status quo G is fixed,

[7] In addition, for any given veto belief m_i, there may be more than one equilibrium to G, under the m_i-revised beliefs. This does not present a problem, but may expand the set of ratifiable mechanisms.

[8] One could define an even stronger notion of ratifiability by considering credible veto beliefs following a veto by a coalition of more than one player and requiring condition (ii) in the above definition to hold for all of these more general kinds of veto beliefs (and all members of the veto coalition) as well. That stronger definition would not change the analysis of the cartel example in the next section, nor the general results about the existence of secure games in Section 5.

[9] Matthews *et al.* (1991) provide an interesting analysis of pre-play communication in Bayesian games that is relevant here. As in Farrell (1985), their pre-play communication is cheap talk (it does not affect payoffs directly) but they allow a richer structure. In particular, different deviant types may send different messages. The communication in our model is quite different, since it stems from the binary decisions to ratify. Communication in our model is not cheap talk and naturally splits the set of types into two subsets, those that vetoed and those that ratified the mechanism.

participation constraints are sometimes strengthened by learning from disagreement and sometimes weakened depending on the proposed mechanism d.

Two firms produce in an industry. Each firm's marginal cost c_i is private information and drawn independently from the uniform distribution on [0, 1]. Inverse demand is $p(q_1, q_2) = 1 - q_1 - q_2$, where q_i is the production of firm i. Both firms seek to maximize their interim profit. The status quo G is Bayesian Cournot competition: each firm i chooses q_i to maximize its expected profit given its cost c_i and its belief about j's production cost, c_j Since both the type and decision spaces are continuous, we extend our definitions in the natural way. Moreover, because of the continuous type space, randomization in the decision to ratify will prove unnecessary, so we can work directly with credible veto sets V_i, rather than credible veto beliefs m_i.

A mechanism d specifies how much each firm produces and how the industry revenue is divided between the firms as a function of their reported costs. We consider two different mechanisms. The first, joint monopoly, illustrates that learning from disagreement can strengthen the participation constraints—although joint monopoly is attainable without learning, it is not ratifiable. The second, a minimum quantity restriction, demonstrates that learning from disagreement can have the opposite affect—although no minimum quantity restriction is individually rational without learning, it is ratifiable.

3.1. *Joint-Monopoly Mechanisms*

Let d be the joint monopoly outcome: the lowest-cost firm produces the monopoly output and the other produces nothing with the revenue divided in such a way that d is incentive compatible. Using the standard mechanism design approach, one can show that it is possible to split up the revenue in such a way as to satisfy interim individual rationality assuming nothing is learned from disagreement (that is, the status quo of Cournot competition is played with the prior beliefs). In this mechanism without learning, the worst-off type, $\hat{c} = 0.2200$, expects to get 0.0999 in profits from the joint-monopoly mechanism and only 0.093 in the Cournot game.

But is the joint-monopoly mechanism ratifiable against Cournot competition? To answer this question, we need to ask: What should firm j think (and consequently how much should it produce) if firm i vetoes the monopoly mechanism? Does there exist a set of types V_i (a credible veto set) that can make the following speech:

> "I voted against this mechanism because my type is in V_i. If you believe me and I am telling you the truth, then my payoff in the Cournot game is better than what I get in the mechanism. Moreover, if my type is not in V_i, I would get a strictly higher payoff from the mechanism than from the Cournot game in which you believe my type is in V_i, and so I would not want to vote against the mechanism. Hence, you should believe me."

If such a credible veto set exists (and there is no other veto set satisfying (ii) of the definition of ratifiability), then the monopoly mechanism is not ratifiable. This turns out to be the case, as we demonstrate below.

The intuition for why a credible veto set exists here is straightforward. In the monopoly mechanism, high-cost firms gain a great deal by participating, whereas low-cost firms gain little. By vetoing the monopoly mechanism, a firm sends a credible signal that it has a relatively low cost. Since a firm with a low cost produces a relatively large amount in the Cournot status quo, the other firm will respond optimally by producing less in G than it would with its prior beliefs. This reduction of output by the other firm increases the profit to the vetoing firm in the status quo—enough, in fact, to make a vetoer with sufficiently low cost prefer the Cournot outcome to the monopoly mechanism. A low-cost firm's signal is credible, since a high-cost firm still does better in the monopoly mechanism, despite the improved status quo that results from vetoing.

PROPOSITION 2. *The monopoly mechanism is not ratifiable against the status quo of Cournot competition.*

The idea behind the proof is that a mechanism is ratifiable only if the ratifier's expected output is sufficiently high in the status quo mechanism following a veto. As the ratifier's expected output increases in the Cournot game, the vetoer's output in the Cournot game declines. In fact, it can be shown that for any beliefs following a veto, the subsequent equilibrium in the Cournot game will have the vetoer produce if and only if its cost is less than some amount. For future reference, denote such a critical cost level for the vetoer by $c_u(V)$, where V denotes beliefs the ratifier has about the veto set. We then show that whenever V is such that $c_u(V) \leq 0.8515$ then the vetoer will be worse off in the subsequent Cournot game than it would be in the monopoly mechanism. Finally, we show that the unique credible veto set, V, is the interval [0, 0.444] which results in a value of $c_u(V) = 0.89 > 0.8515$. The details are in the Appendix.

Learning in this case has the effect of strengthening the participation constraints. If a veto resulted in passive inferences (i.e. no updating), then the monopoly mechanism is individually rational relative to the status quo alternative of Cournot competition. But if the inferences are required to satisfy our credibility conditions, then the monopoly mechanism will not be unanimously ratified by all types of all players.

3.2. *Minimum Quantity Restrictions*

The effect of our credibility requirement is not always to strengthen the participation constraints. We demonstrate this in the context of the same duopoly example, but with a different proposed mechanism. In particular, let d be the decision rule associated with the (unique) equilibrium outcome from Cournot competition with a minimum quantity restriction Q (i.e., if a firm decides to produce, it must produce at least Q).

PROPOSITION 3. *For Q sufficiently small, Cournot competition with a minimum quantity restriction of Q is ratifiable against the status quo alternative with no quantity restriction, but for all $Q > 0$ such a mechanism is not individually rational with passive updating.*

The intuition behind this example is straightforward. A high-cost firm that would produce an amount much less than Q in the status quo mechanism and receive a small profit finds it unprofitable to produce with the minimum quantity restriction, and so gets zero. Such a firm would veto the minimum quantity restriction in favor of the status quo if the status quo is played with the prior beliefs. But it is credible for the ratifier to infer that the vetoer has a high cost as a result of the veto, which makes the veto unprofitable. Hence, with learning no type wants to veto.

4. ALTERNATIVE FORMULATIONS OF RATIFIABILITY

The issue of ratifiability is closely intertwined with the issue of defining beliefs "off the equilibrium path" in a game of incomplete information. Our definition of ratifiability challenges the notion that any beliefs, and in particular passive beliefs, can follow an unexpected veto.

In the last several years, a considerable literature on "equilibrium refinements" has built up around precisely this problem of specifying plausible beliefs off the equilibrium path. We now compare our choice of refinement and its implications for ratifiability to some alternative refinements. The purpose of this section is to motivate and clarify why we settled on a definition of ratifiability based on the refinement of Grossman and Perry (1986). The main problem with the other refinements is that they generally violate a natural "rational expectations" requirement that ratifiability satisfies: those types that are believed to have vetoed are precisely the types that benefit from vetoing if everyone believes that they are the types to veto.

To save space, formal definitions are omitted and instead we show how these alternative refinements would apply to the duopoly problem analyzed in the previous section, with costs uniformly distributed on [0, 1] and G given by Cournot competition. As in the first part of the last section, let d be the monopoly mechanism. In this section we are interested in determining whether the monopoly mechanism is unanimously ratified relative to the Cournot status quo, when we apply alternative restrictions on beliefs following a veto. In particular, we investigate three refinements: the intuitive criterion, divinity, and universal divinity.

Before proceeding, it is useful to restate one property of ratifiable mechanisms in the context of our duopoly example. Recall that in the proof of Proposition 2 we showed that d is ratifiable relative to G, so long as the veto set V is such that the highest cost vetoer to produce is sufficiently small ($c_u(V) \leq 0.8515$), or equivalently if the ratifier's expected output is sufficiently high in G. In fact, this is true regardless of which refinement concept we use to place restrictions on veto sets. Hence, d is ratifiable against G relative to a particular refinement if the refinement allows a veto set V such that $c_u(V) \leq 0.8515$.

4.1. *The Intuitive Criterion*

The *intuitive criterion* (Cho and Kreps, 1987) requires that beliefs be concentrated on those types for whom there exists some belief such that if the ratifying firm inferred those beliefs from a veto, then this type of firm would wish to veto. In other words, no weight can be put on types for whom the deviation of vetoing is "bad" in the sense that the vetoer surely loses from the deviation (prefers the mechanism to the status quo) regardless of what belief the veto induced in the other firm.

This is a much weaker requirement than the one we proposed in the previous section. If a mechanism is ratifiable relative to the status quo, then there will always be a sequential equilibrium of the ratification game where everyone ratifies the mechanism and beliefs satisfy the intuitive criterion. In fact, for this example, there are beliefs satisfying the intuitive criterion that support the monopoly mechanism. This is demonstrated below.

We begin by calculating the *believable set*, the set of types for which there exists some belief that makes vetoing profitable. Regardless of type, the belief that makes vetoing the most profitable is $c = 0$, because it leads to the smallest production by the other firm in the status quo. Therefore, type c is believable if c does at least as well by vetoing the mechanism if the ratifying firm infers that the vetoer's cost is 0. With beliefs $V = \{0\}$, $c_u = 0.93$. Beliefs must be concentrated on types for whom $U(c,\{0\}) \leq 0$, which implies that the believable set is [0, 0.562].

The belief for the ratifier satisfying the intuitive criterion that is most apt to support the monopoly mechanism is the most *optimistic* belief; namely, for the ratifier to infer that the vetoers cost is the largest believable type $c = 0.562$. It is easy to verify that such a belief supports the monopoly mechanism: all types prefer the monopoly outcome to the status quo when by vetoing the mechanism the vetoer reveals that its cost is 0.562. For beliefs $V = \{0.562\}$, $c_u < 0.8515$, so the monopoly mechanism is ratifiable under the intuitive criterion.

The intuitive criterion is a weak refinement in this application, and similar applications with a continuum of types, because of the extreme beliefs that it allows. The believable set is defined by the ratifying firm making the *most pessimistic* inference ($c = 0$), but then the individual rationality constraint is determined by making the *most optimistic* inference ($c = 0.562$). The intuitive criterion allows beliefs that are far from consistent in a rational expectations sense.

4.2. *Divinity*

The *divinity* refinement (Banks and Sobel, 1987) goes one step further than the intuitive criterion by imposing a monotonicity condition on beliefs in the believable set. For any two believable types, t and t', with the property that the set of beliefs under which t' prefers to veto strictly contains the set of beliefs under which t' prefers to veto, devine beliefs must have a higher likelihood ratio of t to t' than the prior likelihood ratio between the two types. As in the intuitive criterion, we must find the most optimistic belief possible, but subject to the likelihood ratio constraint. It is easy to show that the most optimistic divine belief is simply the truncated prior on the believable set. In our example, this reduces to a uniform posterior on the believable set. With this belief, a veto of the mechanism will result in Cournot competition in which the vetoer believes that the ratifying firm's cost in uniformly distributed on [0, 1] and the ratifier believes that the vetoing firm's cost is uniform on [0, 0.562]. With $V = [0, 0.562]$, $c_u = 0.878 > 0.8515$, so the monopoly mechanism is not ratifiable with divine beliefs. In our example, divinity is strong enough to eliminate the monopoly outcome.

4.3. *Universal Divinity*

The stronger refinement of *universal divinity* (Banks and Sobel, 1987) is a strengthening of divinity that requires the likelihood ratio condition to hold relative to *any* prior, not just the original prior. This condition essentially requires that a belief places zero probability not only on types outside the believable set, but also on most other types as well. Which types must receive zero probability is determined in the following way. For every belief p concentrated on the believable set, let $C(p)$ denote the set of types who are at least as well off vetoing the monopoly mechanism and playing the Cournot game in which the ratifier has these beliefs about the vetoer's type and the vetoer has the original prior about the ratifier, and let $C(p)$ denote the set of types who strictly prefer vetoing the monopoly mechanism when the ratifier has beliefs p in the status quo. A type c must receive zero probability in a universally divine belief if, for every belief p in the believable set, if $c \in C(p)$ then there exists another type $c' \in C(p)$.

If the payoff structure of a game possesses a monotonic structure in types, universally devine beliefs are concentrated at one point, corresponding to the type which, at least for some beliefs p in the believable set, is the unique member of $C(p)$ and, for all p' concentrated in the believable set, is in $C(p')$ whenever any other type is in $C(p')$. More generally, this criterion reduces to a set of types, which can be called the *universally divine set.* This set will include (at least) every type for which there exists some belief p concentrated in the believable set with the property that that type is the unique element of $C(p)$ and will exclude every element which is never a unique element of $C(p)$ for any p concentrated in the believable set.

In our example, universally divine beliefs are not difficult to calculate, since there is a unique type in the universally divine set; namely, $\hat{c} = 1/2 - (c_u - 3/4)^{1/2}$ where $c_u = 0.8515$, so $\hat{c} = 0.1814$. (A belief that yields $c_u =$

0.8515 is $V = \{0.3912\}$.) It is easy to show that with beliefs concentrated on 0.1814, a nonempty range of types strictly prefer to veto. Therefore, the monopoly mechanism is not universally divine.

All of the refinements we consider, except the intuitive criterion, suggest that the monopoly mechanism is implausible, because of difficulties in getting the monopoly mechanism ratified. However, we find our concept of ratifiability more appealing than the other refinements, because the others are not consistent in a rational expectations sense. The set of types who benefit from vetoing is different from the set of types who are assumed to have vetoed the mechanism.

5. SECURE MECHANISMS

So far we have taken the status quo mechanism G to be fixed, which is consistent with the notion that the mechanism designer cannot alter the status quo. However, in some situations the mechanism designer instead might wish to identify an institution that can stand up against alternative i institutions. The institution can be thought of as a status quo mechanism. It is "secure" if there is no alternative mechanism that can be strongly ratified in favor of the institution. The definition of security uses the strong definition of ratifiable to make it more comparable with durability.

DEFINITION. A mechanism G is *secure* if every incentive compatible decision rule that can be strongly ratified against G is an equilibrium outcome of G under the prior beliefs.

DEFINITION. An incentive compatible decision rule d is *secure* if there exists a secure mechanism G with an equilibrium s where d is the equilibrium outcome in G under s.

5.1. *Security and Durability*

We next compare security to a related concept, durability (Holmström and Myerson, 1983), which also attempts to formalize the idea of a mechanism being invulnerable to proposals of alternative mechanisms. As Crawford (1985) has discussed at length, a thorough analysis of mechanism design requires a careful study of the variety of procedures for choosing outcomes a group of privately informed agents might use.

Durability suggests the following story. There is a direct revelation mechanism d (the status quo) and an alternative decision rule gam. Players each receive private information and then vote between gam and d in the first stage. In the second stage, they play either g or d depending on whether g is unanimously approved in the voting stage. Prior to this second stage the players are not told the details of the vote, only whether g or d is being played. Consider all trembling-hand perfect equilibria of this two-stage game. g is said to be *rejectable against* d if there is at least one equilibrium in which there is no equilibrium path where g is unanimously approved over d. In other words, there is an equilibrium in which, for every realization of types, at least one player rejects g. If every g can be rejected, then d is *durable.*

A *secure mechanism is* also defined as a status quo that cannot be unanimously voted down vis-a-vis some other incentive compatible decision rule, but there are some critical differences between security and durability. First, our status quo mechanism G is not necessarily a direct mechanism, as in durability, and we only consider direct mechanisms as proposals, while durability considers arbitrary mechanisms as proposals. Second, security differs from durability in its definition of ratifiability (as opposed to rejectability). With security, g is ratifiable over G if there exists an equilibrium in which g is unanimously approved over G *at every profile of types.* Thus an alternative mechanism g is "rejected" according to the security definition (i.e. not ratified) if there is no equilibrium to the two-stage game in which g is unanimously approved over G along *every* equilibrium path (i.e. every profile of types). That is, g is rejectable against G if, for every equilibrium of the two stage ratification game, there exists a profile of types in which g is not unanimously approved over G.

Therefore, the difference between secure mechanisms and durable mechanisms is as follows. In the security definition, a rejection of a proposed mechanism can be made by a single type of a single player; in the durability definition, a rejection of a mechanism requires at least one type for every profile of types. Or, conversely, d only endures g if g is rejected for *every* profile of types, while d is secure against g if g is rejected at any *one* profile of types. Therefore, it would seem that if d endures g then d is secure against g, or alternatively if g is ratifiable against d then g is not rejectable against d. However, the relationship between durability and security is not that simple since

security permits more general game forms than durability, which considers only direct mechanisms. The identification of general conditions under which this relationship holds is an open question.

Part of the difficulty in identifying such a set of conditions is that a more subtle difference between the two definitions arises that has to do with passive beliefs and zero-probability events. It is explicitly assumed in security that if g is ratified it is played with the original priors (i.e. passive beliefs). On the contrary, for durability, passive beliefs hold whenever g is *not* ratified, i.e. when it is rejected in favor of the status quo d. Thus durability allows only for learning from *agreement* to the alternative, while security considers only learning from disagreement in favor of the status quo.

This distinction reflects somewhat different implicit assumptions about the temporal sequence of commitments by the players to the mechanism. Holmström and Myerson (1983) are trying to capture (at least in part) the idea of one type of one player proposing the alternative g at the interim stage, after d has been set up at the ex ante stage. The goal of the designer is to create an efficient d that is invulnerable to such interim proposals.[10] Security views the proposal g as being made at the ex ante stage, but the *vote* between g and G being made at the interim stage. This limits the possibilities for defeating G with an unrejectable proposal at the interim stage.

5.2. *An Example*

We now consider an example due to Holmström and Myerson (1983), which illustrates the difference between durability and security. Two players have independent types and private values. Each player is one of two equally likely types: player 1 is either type la or type lb and player 2 is either type 2a or 2b. There are three possible decisions $\{A, B, C\}$. The players payoffs depend on their types and which decision is adopted, as tabulated below:

	1a	1b	2a	2b
A	2	0	2	2
B	1	4	1	1
C	0	9	0	–8

The incentive compatible decision rule that uniquely maximizes the equally weighted sum of expected payoffs is

	2a	2b
1a	*A*	*B*
1b	*C*	*B*

Call this decision rule *ABCB.* Holmström and Myerson point out that this is *not* durable since the allocation rule *AAAA* defined by

	2a	2b
1a	*A*	*A*
1b	*A*	*A*

will not be rejected for the type profiles (la, 2a) and (la, 2b).

It is easily verified, however, that in the context of ratifiability, player 1 would veto *AAAA* in favor of *ABCB* if his type were lb. Therefore, *AAAA* is *not* ratifiable against *ABCB.* Interestingly, however, this does not mean that *ABCB is* a secure mechanism. Consider the following decision rule, *AABB:*

[10] One must be careful not to interpret this story too literally, since the actual voting game considered by Holmström and Myerson does not include the proposal-making stage. This circumvents a number of subtle issues about information leakage through proposal making.

	2a	2b
1a	*A*	*A*
1b	*B*	*B*

According to *AABB,* player 1 may unilaterally choose between *A* and *B.* This allocation rule is ratifiable against the direct mechanism *ABCB even though it is not ratifiable if vetoing generates passive beliefs.* The reason is that player 2 would never veto; in fact, it is easy to show that there are no credible veto beliefs for player 2. Moreover, the only credible veto belief for player 1 is $m\{1b\} = 1$ (i.e., the only credible veto set is {lb}). Therefore, if lb vetoes, the outcome of *AABB* under the updated beliefs will have player 2 announcing 2b, regardless of his type, and player 1 will announce lb. This generates the same utility to lb that he would have had by not vetoing, since the outcome is always *B* when he is type lb. Clearly player la is strictly worse off vetoing. Therefore, *AABB* is ratifiable against *ABCB.* The intuition is that lb types cannot veto *AABB* without giving away their identity, and if they give away their identity there is no advantage to vetoing.

The following example shows that there exist mechanisms that are not durable, but are secure. It is a variation on the previous example where types are not independent. Specifically, the probability distribution of types is

	2a	2b
1a	0.5	0.1
1b	0.2	0.2

Consider the following allocation rule *ABBA:*

	2a	2b
1a	*A*	*B*
1b	*B*	*A*

It is easy to verify that *ABBA is* incentive compatible and interim efficient. But *ABBA* is *not* durable, since the *AAAA* allocation rule is incentive compatible and not rejectable against *ABBA.* Type la will not vote to reject, and neither of player 2's types will vote to reject. However, this alternative is not ratifiable, since $m\{1b\} = 1$ is a credible veto belief for player 1. This type expects an interim utility of 0 under *AAAA,* but earns an expected utility higher than that under *ABBA,* even when vetoing reveals his type to player 2. The new equilibrium in the status quo mechanism played under the beliefs that player 1 is type lb involves mixing by both players and gives both types of player 1 a probability of decision *A* that is strictly between 0 and 1. This gives type la a lower utility than *AAAA,* and gives lb a higher utility than *AAAA,* thus verifying {lb} as a credible veto set for player 1.

To establish that *ABBA* is secure, we need to check every other incentive compatible decision rule and verify that there is a credible veto belief in each case. This tedious search is simplified by establishing the following general property of secure allocation rules, which is also used in the next section to prove an existence result.

LEMMA 1. *Consider a status quo mechanism defined by the direct mechanism associated with allocation rule d, and an alternative allocation rule d'. Then d' is not strongly ratifiable against d if there exists some player i and some type t_i such that the interim utility of d is strictly greater than the interim utility of d' for t_i, and the interim utility of d is greater than or equal to the interim utility of d' for all $t_i' \in T_i$.*

Proof. The proof consists of showing that the prior p_i is a credible veto belief with corresponding credible veto set T_i. Since the interim utility of d is greater than the interim utility of d' for all $t_i' \in T_i$, we get $U_i(t_i', s_0(p_i)) \geq 0$ for all $t_i' \in T_i$, where $s_0(p_i)$ is the truth-telling equilibrium of the direct mechanism defined by d with the prior belief p_i. Since the interim utility of d is strictly greater than the interim utility of d' for t_i, we get $U_i(t_i, s_0(pi)) > 0$. Therefore, p_i is a credible veto belief for player i, so d' is not strongly ratifiable against d.

In this example, it is easy to see that any alternative incentive compatible allocation rule other than *AAAA* makes at least one type of player 2 strictly worse off than *ABBA* and makes no type of player 2 better off. The security of *ABBA* follows immediately.

5.3. *Existence of Interim Efficient Secure Mechanisms*

Holmström and Myerson (1983) establish existence of interim efficient durable mechanisms quite generally, by showing that there exist durable mechanisms that maximize one player's interim utility subject to what are effectively durability constraints, applying arguments from Myerson (1983). Those techniques cannot be applied directly to prove existence of secure mechanisms, since security does not assume passive updating in the event of the rejection of an alternative mechanism.

However, one can establish existence for a class that includes private values environments. (By private values, we mean that a player's utility function does not depend on the types of the other players.) Specifically, we need an assumption used in the Bayesian implementation literature (Jackson, 1991; Matsushima, 1990).

DEFINITION. An individual i has *known best elements* if, for all $d \in D$, $t \in T$, $t'_{-i} \in T_{-i}$,

$$u_i(d,t) \geq u_i(d',t) \quad \text{for all} \quad d' \in D \Rightarrow u_i(d,t_i,t'_{-i}) \quad \text{for all} \quad d' \in D.$$

Thus, i has known best elements if, for every t_i, i's top ranked alternatives do not depend on t_{-i}. If the known best elements for a player are unique for every type of that player, then we say the player has *unique known best elements.*

THEOREM 1. *If at least one player has unique know best elements, then there exists an interim efficient and secure mechanism.*

Proof. Let i have unique known best elements, denoted $d_i^*(t_i)$. Then $d(t) = d_i^*(t_i)$ is interim efficient. From Lemma 1 it is secure.

THEOREM 2. *Suppose the set D of possible decisions is finite. every player has private values, then there exists an interim efficient and secure mechanism.*

Proof. Consider the following mechanism. Player 1 moves first and selects a nonempty subset $D_1 \subset D$. Then player 2 moves and selects a nonempty subset $D_2 \subset D_1$. This continues with player i selecting a nonempty subset $D_i \subset D_{i-1}$ at stage i. Finally, player n selects a single element $d \in D_{n-1}$, and the game ends with the outcome being d. For each i and $t_i \in T_i$, and nonempty $C \subset D$, denote by $B_i(t_i, C)$ that type of that player's set of known best elements restricted to C. It is a sequential equilibrium for each type of each player to choose, at each stage, the set $B_i(t_i, D_{i-1})$ and for player n to choose any element of $B_n(t_n, D_{n-1})$. Because of private values, the $B_i(\cdot)$ sets do not depend on player i's beliefs about the other players' types, and this equilibrium generates an efficient allocation rule. Therefore, this is a sequential equilibrium following any veto of any mechanism by any type of any player.

6. CONCLUSION

A basic tenet of most studies in mechanism design is that the parties voluntarily decide to participate in a proposed mechanism. Indeed, if participation constraints are ignored, then it typically is possible to overcome incentive problems caused by informational differences. Our purpose in this paper has been to take a closer look at participation constraints and a party's decision to participate. Central to this goal is specifying what happens if the parties fail to agree to participate in a proposed mechanism. We consider situations where the appropriate "status quo" is a noncooperative game, rather than some constant payoff. In this case, the status quo payoffs may depend on inferences based on the parties' participation decisions. Participation constraints, then, depend on the beliefs the parties hold following a veto of the proposed mechanism. This paper has proposed and illustrated a concept of ratifiability that takes into account in a consistent way the dependence of beliefs on inferences from a veto. The concept of ratifiability leads naturally to the notion of a secure mechanism, against which any alternative mechanism

will fail to be ratified. The existence of interim efficient, secure mechanisms may require some degree of private values in the environment.

APPENDIX

Proof of Proposition 2. We first calculate the unique credible veto set V. Since the two firms are symmetric ex ante, we can drop the subscripts i and j in what follows. Furthermore, it is easy to show that for any veto set V there is a unique equilibrium s in G, so we can indicate U_i^G 's dependence on $s(V)$ simply as V. We need to calculate the interim payoffs from the monopoly mechanism $U_i^d(c)$ and from the status quo $U_i^G(c,V)$ for every c and every V. The first is easy to compute from a formula in Cramton and Palfrey (1990):

$$U_i^d(c) = [1/8 + (1-c)^3]/6.$$

The second function is harder to derive because the strategies for the two firms differ, since each firm has different beliefs about the other.

Denote the vetoer's strategy by $q_u(c, V)$ and the ratifier's strategy by $q_r(c, V)$. In either case, the strategy depends on a single cost level at which the firm is indifferent between producing and not. Let $c_u(V)$ and $c_r(V)$ be the indifference cost levels for the vetoer and ratifier respectively. Then

$$q_u(c,V) = \begin{cases} (c_u(V) - c)/2 & \text{if } c < c_u(V) \\ 0 & \text{if } c \geq c_u(V) \end{cases}$$

$$q_r(c,V) = \begin{cases} (c_r(V) - c)/2 & \text{if } c < cr(V) \\ 0 & \text{if } c \geq cr(V) \end{cases}$$

Let $Q_u(V)$ and $Q_r(V)$ be the expected output of the vetoer and the ratifier, respectively. Then the cost levels at which a firm is indifferent between producing and not producing are

$$c_u(V) = 1 - Q_r(V), \tag{Cv}$$

and

$$c_r(V) = 1 - Q_u(V), \tag{Cr}$$

The vetoer calculates the ratifier's expected output with the prior belief, since the veto is a unilateral deviation. Thus,

$$Qr(V) = \int_0^1 qr(c,V)dc = \tfrac{1}{4}c_r(V)^2. \tag{Qr}$$

Suppose V is an interval $[\ell, h]$. Then the ratifier calculates the vetoer's expected output to be

$$Q_u(V) = \int_\ell^h \frac{q_u(c,V)}{h-\ell}dc = \begin{cases} 0 & \text{if } c_u(V) \leq \ell \\ \dfrac{(c_u(V)-\ell)^2}{4(h-\ell)} & \text{if } \ell < c_u(V) < h \\ \tfrac{1}{2}(c_u(V) - (\ell+h)/2) & \text{if } c_u(V) \geq h\,. \end{cases} \tag{Qv}$$

The equilibrium in the status quo with beliefs $V = [\ell, h]$ *is* found by solving equations (Cr), (Cv), (Qr), and (Qv) for $c_r(V)$, $c_u(V)$, $Q_r(V)$, and $Q_u(V)$. The vetoer's payoff in the status quo then is

$$U_i^G(c,V) = q_u(c,V)^2\,.$$

Let $U(c, V) = U(c,V) = U_i^d(c) - U_i^G(c,V)$. For a veto set to be credible, it must be that

$$U(c,V) \le 0 \text{ for all } c \in V \qquad \text{and} \qquad U(c,V) \ge 0 \text{ for all } c \notin V. \tag{CV}$$

First, note that $U(c, V) > 0$ for all $c \ge c_u(V)$, so V cannot contain $c \ge c_u(V)$. For $c < c_u(V)$, $U(c, V)$ is given by the cubic equation

$$U(c, V) = (\tfrac{1}{8} + (1 - c)^3)/6 - (c_u(V) - c)^2/4.$$

Thus, for any V, $U(c, V)$ is parameterized by the single cut-off level $c_u \in [0.75, 0.93]$ ($c_u = 0.75$ for $V = \{1\}$ and $c_u = 0.93$ for $V = \{0\}$). It is easy to show that $U(c, V)$ is (1) continuous, (2) decreasing for c near 0, (3) has a minimum at $c = \hat{c} = 1/2 - (c_u - 3/4)^{1/2}$, (4) has a point of inflection at $c = 1/2$, and (5) has a maximum at $c = 1/2 + (c_u - 3/4)^{1/2}$. Hence, d is individually rational relative to G if and only if V is such that c_u is sufficiently small. Calculation shows that this critical cutoff level is $c_u = 0.8515$. Moreover, the only candidate for a credible veto set is an interval $[\ell, h]$ around $\hat{c}$. One such veto set takes the form $V = [0, h]$, where $h < c_u(V)$ *is* such that $U_i^d(h) = U_i^G(h,V)$*;* that is, h solves $(1/8 + (1 - h)^3)/6 = ((c_u(V) - h)/2)^2$ and $c_u(V)$ solves (Cr), (Cv), (Qr), and (Qv). These equations are satisfied when $h = 0.444$. In this case, $c_r(V) = 0.67$, $c_u(V) = 0.89$, $Q_r(V) = 0.111$, $Q_u(V) = 0.333$, $U_i^d(h) = U_i^G(h,V) = 0.0496$, and (CV) is satisfied, since $U(0, V) < 0$. This veto set is unique, since raising h decreases c_u, so $U(h, V) > 0$ and h will no longer want to veto.

Proof of Proposition 3. We must show that for small Q there exists a credible set V which satisfies (ii) of the definition of ratifiability. Under d, the two firms simultaneously select output levels q_1 and q_2 subject to the constraint that if $q_i > 0$ then $q_i \ge Q$. The status quo G is the same as before: Cournot competition without quantity restrictions. A unique equilibrium is characterized for each Q by two cutoff levels, ℓ and h, with $\ell \le h$. Each firm produces

$$q(c) = \begin{cases} 0 & \text{if } > h \\ Q & \text{if } \le c \le h \\ Q + \frac{1}{2}(\ell - c) & \text{if } c < \ell, \end{cases}$$

where the cutoff levels are determined from the equations

$$Q = (h - \ell)/2 \qquad \text{and} \qquad \ell^2/4 = 1 - (1 + Q)h.$$

If $Q = 0$, then this mechanism is simply Cournot competition with the prior beliefs, and $\ell = h = 0.83$. As Q increases, both ℓ and h decrease. This means that, relative to the status quo with the original priors, no such mechanism is individually rational with passive beliefs: a firm with costs slightly below 0.83 makes 0 in the mechanism but would make some small positive profit under Cournot competition.

This ignores the inferences that the other firm would make if such a firm were to veto. Intuition suggests that a vetoing firm will be suspected of having high costs, thereby destroying any benefits a high cost firm could get from vetoing the minimum-quantity mechanism. In fact, this intuition is correct for sufficiently small values of Q. For small Q, one credible veto set is simply the highest cost type, $c = 1$. Such an inference from a veto will lead a ratifying firm to produce much more in the status quo than it would under its original priors which makes all types (weakly) prefer the mechanism to the status quo. A $c = 1$ type earns zero profits in either case, so vetoing or ratifying by such a type can be rationalized. Thus, a minimum quantity restriction is a ratifiable mechanism, even though it is not individually rational with passive updating.

ACKNOWLEDGMENTS

We thank Jeff Banks, Steve Matthews, Preston McAfee, Joel Sobel, numerous seminar participants, and two referees for valuable comments. We are grateful to the National Science Foundation for financial support. The first author thanks the Hoover institution of Stanford University for a most enjoyable year as a National Fellow.

REFERENCES

AUSUBEL, L. M., AND DENECKERE, R. J. (1989). A Direct Mechanism Characterization of Sequential Bargaining With One-Sided Incomplete Information," J. *Econ. Theory* **48,** 18-46.

BANKS, J. S., AND SOBEL, J (1987). "Equilibrium Selection in Signaling Games " *Econometrica* **55,** 647-662.

CAILLAUD, B., AND HERMALIN, B. (1989). The Role of Outside Considerations in the Design of Compensation Schemes," working paper. Berkeley: University of California.

CHO, I.-K., AND KREPS, D M. (1987). "Signaling Games and Stable Equilibria." *Quart.* J. *Econ.* **102,** 179-222.

CRAMTON, P. C. (1985). Sequential Bargaining Mechanisms," in *Game Theoretic Models of Bargaining* (A. Roth, Ed.). Cambridge: Cambridge University Press.

CRAMTON, P. C., AND PALFREY, T. R. (1990). Cartel Enforcement with Uncertainty about Costs," *Int. Econ. Rev.* **31,** 17-47.

CRAWFORD, V. P. (1985) Efficient and Durable Decision Rules: A Reformulation," *Econometrica* **53,** 817-836.

FARRELL, J. (1985). Communication and Nash Equilibrium," working paper. Cambridge, MA: M.I.T.

GREEN, J. R., AND LAFFONT, J.-J. (1985). "Implementation through a Sequential Unanimity Game," working paper. Cambridge, MA: Harvard University.

GREEN, J. R., AND LAFFONT, J.-J. (1987). Posterior Implementability in a Two-Person Decision Problem, *Econometrica* **55,** 69-94.

GROSSMAN, S. J., AND PERRY, M. (1986). Perfect Sequential Equilibrium, J. *Econ. Theory***39,** 97-119.

HOLMSTRÖM, B. AND MYERSON, R. B. (1983). "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica* **51,** 1799-1819.

JACKSON M. O. (1991). Bayesian Implementation," *Econometrica* **59,** 461-478.

LEGROS, P. (1991). Stable Allocations," working paper. Ithaca, NY: Cornell University.

MASKIN, E., AND TIROLE, J. (1990). "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values," *Econometrica* **58,** 379-410.

MASKIN, E., AND TIROLE, J. (1992). "The Principal-Agent Relationship with an Informed Principal. II. Common Values," *Econometrica* **60,** 1-42.

MATTHEWS, S., OKUNO-FUJIWARA, M., AND POSTLEWAITE, A. (1991). Refining Cheap-Talk Equilibria, J. *Econ. Theory* **55,** 247-273.

MATUSHIMA, H. (1990) Characterization of Full Bayesian Implementation," working paper. Stanford, CA: Stanford University.

MAYERSON R. B (1983) Mechanism Design by an Informed Principal," *Econometrica* **51,** 1767-1797.

MYERSON, R. B. (1985). "Bayesian Equilibrium and Incentive Compatibility: An Introduction." *Social Goals and Social Organization: Essays in Honor of Elisha Pazner* (*L.* Hurwicz, D. Schmeidler, and H. Sonnenschein, Eds.), pp. 229-259. Cambridge: Cambridge University Press.

MYERSON, R. B., AND SATTERTHWAITE, M. A. (1983). Efficient Mechanisms for Bilateral Trading,' J. *Econ. Theory* **28,** 265-281.

ROBERTS, K. (1985) Cartel Behaviour and Adverse Selection," *J. Ind. Econ.* 33-45.

SPIER, K. E (1988). Efficient Mechanisms for Pretrial Bargaining," working paper. Cambridge, MA: M.I.T.

SPULBER, D F. (1989). Contingent Damages and Settlement Bargaining," working paper. Los Angeles: USC.