COMMITMENT AND TIME CONSISTENCY:
A GAME-THEORETIC DISCUSSION

Daniel B. Klein
Department of Economics
University of California
Irvine, California 92717

Brendan O'Flaherty
Department of Economics
Columbia University
New York, New York 10027

Revised: August 1993

KEYWORDS: time inconsistency, commitment dominance, policy set, sequential irrationality, concatenated strategy, S-game.

JEL classification numbers: C7

We have benefitted from comments by Brian Hillier, David Lilien, Stergios Skaperdas, and participants at workshops at Stanford, Cal Tech, UCSB, UCLA, UCB, UCI, and USC.

Commitment and Time Consistency: A Game-Theoretic Discussion

by

Daniel Klein and Brendan O'Flaherty

Abstract

The paper offers a game-theoretic framework for discussing commitment and time consistency. We show when a commitment (or just the appearance of one) is valuable, how valuable it is, and whether the commitment is time consistent.

We formulate restrictions on the set of eligible commitments. These restrictions correspond to the "open-loop" programming format, and we call the restrictions "policy formation."

We give a game-theoretic depiction of Fischer's savings-taxation model of time inconsistency. This depiction shows in a game tree the essential lessons of the model.

The emphasis is on definition, examples, and interpretation. The paper contains no formal propositions. Our goal is to show that the concepts developed are useful for understanding commitment and time consistency, for modeling situations with these characteristics, and for achieving general results.

1. PRELIMINARIES

In many strategic scenarios being able to make a commitment helps a player. Indeed, in the scenario-rich work of Thomas Schelling, commitment emerges as one of the central principles of strategy. In Schelling's discussion (1960) the actor in need of commitment is likely to be a kidnapper, a terrorist, or a military chief. In the public finance and macroeconomics literature the actor is usually benevolent governments.

In this paper we offer a framework for discussing commitment and time consistency. We do not give any account of commitment technology, of *how* a commitment is conveyed. We show when a commitment (or the appearance of one) is valuable, how valuable it is, and whether the commitment is time consistent. Also, we apply the concepts developed by giving a game-theoretic discussion of Fischer's savings-taxation model of time inconsistency (1980). Our rendering of time inconsistency is *not* a new definition; rather it is the standard definition fitted to a more general and more complete framework.

Two Ways of Getting Your Back Scratched: Promises and Threats

One way to get your back scratched it to tell the other guy, "If you scratch my back I'll scratch your back." The situation is depicted in Figure 1. (Disregard for the moment the probabilities shown at v. Suppose that only pure strategies are permitted.) Player 1 makes this promise. She had better make the promise convincing, for otherwise Player 2 will treat it as hot air,

think that Player 1 will not scratch his back, and, therefore, not scratch Player 1's back. Clearly Player 1 values commitment conveyance, for without it no backs are scratched.

[Figure 1 here.]

The promise is time inconsistent because once Player 2 has scratched Player 1's back, Player 1 would like to dissolve any commitment and revert to not scratching. It is precisely this inconsistency that makes the commitment a promise.

A second way to get your back scratched is to tell the other guy, "Scratch my back or else I'll break your back." The situation is depicted in Figure 2. Again, Player 1 needs to make the plan convincing, for otherwise Player 2 will not oblige. Notice that although commitment conveyance is again of value, the plan is time consistent: the ruler (Player 1) does not get to a node where she would like to revise the plan. This plan is a threat.

[Figure 2 here.]

Schelling (1960, 177) pointed out the distinction: " [A] promise is different from a threat. The difference is that a promise is costly when it succeeds, and a threat is costly when it fails. A successful threat is one that is not carried out," whereas a successful (and genuine) promise is carried out. Not all commitments are neatly categorized into pure promise and pure threat. We will see an example that combines elements of both. For a fuller discussion of promises and threats building on Schelling's original ideas and using the framework developed in this paper, see Klein & O'Flaherty (1993).

The Inadequacy of Noncooperative Game Theory to Deal with Commitment

Despite its obvious importance, the idea of commitment has eluded the scenario-free world of noncooperative game theory.[1] The common suggestion is to "write the commitments into the game" and use the noncooperative concepts. But this approach suffers from four shortcomings.

In Figure 3 we take this suggestion for the commitment scenario of Figure 1 ("If you scratch my back I'll scratch your back"). We restrict attention to pure strategies. In this formulation commitments become Player 1's first moves. She can either commit to *quid pro quo* or not commit. Each of these moves then leads to Player 2's binary decision, and Player 2's choice will imply payoffs since Player 1 has already determined her behavior. Naturally the unique perfect equilibrium conforms to the same outcome of the commitment story told for Figure 1 (namely, mutual scratching).

[Figure 3 here.]

The first shortcoming of this traditional formulation is that in the raw structure of the game the commitment interpretation has been lost. Only once a story is told about back scratching, only after labels are put on Player 1's moves, can anything in the model be interpreted as a commitment. The traditional formulation suppresses any hope of a pure theory of commitment. As Schelling pointed out (1960, p156): Once we write the commitments into the game tree "[t]he objects of our study, namely, these tactics together with the communication and enforcement structures that they depend on, and the timing of moves, have all disappeared." Building the commitment into the game tree is acceptable for the telling of *specific stories*, but it fails to give

commitment a theoretical status distinct from game moves. It shuts the door on establishing general results about commitment, since the commitment interpretation remains only as a piece of the incidental story that accompanies the bare machinery. In our formulation commitment notions exist in the bare machinery.

The second shortcoming of the traditional formulation is that there is no easy way to indicate Player 1's desire to undo her apparent commitment once Player 2 has scratched her back. Thus there is no scope for a theory of time inconsistency.

The third shortcoming is that a traditional formulation will complicate the tree acutely. As situations grow more complex, the number of possible commitments grows rapidly. The fourth shortcoming is that the traditional formulation deals inaptly with mixed strategy commitments. In Figure 1, if mixed strategy commitments are permitted, the optimal plan for Player 1 is to elicit back scratching with the least costly promise, namely, the promise that scratching will be returned with only 0.75 probability. Player 1's expected payoff is then $0.75*4 + 0.25*5 = 4.25$.[2] To deal with such mixed strategy plans in the traditional formulation would require additional branches in the tree in Figure 3.

S-games

Our approach is to develop a new set of lenses with which to look at certain strategic situations. What we see through these lenses is not a traditional game but what we call an "S-game" (for Schelling and for Stackelberg). An S-game specifies a traditional extensive form game, but it specifies some other

things as well. First it designates one player as the "ruler," who pre-emptively announces her strategy in the game. Second, it specifies determinate responses to the ruler's announcement. True to the Stackelberg tradition, the ruler can convey commitments, whereas the other players ("the public") are confined to playing a subgame perfect equilibrium in the game induced by the ruler's announcement. Third, an S-game specifies certain important restrictions on the set of plans the ruler can announce. (These restrictions correspond to the "open-loop" programming format.)

Two Interpretations of Commitment Conveyance

The announced plan is received by the public with complete credulity. As investigators, we may remain unfixed on whether the public's belief in the announcement is warranted. We do not assume that the ruler's promises and threats are credible (that is, worthy of belief); we assume only that they are credited (or believed).

You can think about our analysis in two ways:

Interpretation 1: The ruler conveys genuine commitments.

Interpretation 2: The ruler conveys phony commitments.

Interpretation 1 says that it is common knowledge that the ruler must truly commit to the announced plan at the start of the game. We then ask: If at any of the ruler's nodes a genie were to pop out of a bottle and allow the ruler to deviate from the announced plan, would she deviate?[3] When the public sees the genie and the ruler's deviation it correctly anticipates

sequentially rational play from the deviation onwards and adjusts its play accordingly. The genie never actually appears.

Interpretation 2 says that the ruler does not really commit to the announced plan but the public thinks she does. The public's belief is mistaken. If the ruler ever deviates from the announced plan, the public then realizes that her hands are not tied and that her behavior for the remainder of the game will be sequentially rational.

Each interpretation has appealing aspects. As already stated, there is no need to settle on one or the other. Because of the dual interpretations, rather than saying that the ruler makes a commitment, we say that the ruler *conveys* a commitment.

* * *

This paper is devoted to definitions, examples, and interpretation. It strives for accessibility. It contains no formal propositions. The concepts developed here have been used elsewhere to achieve general results. (Klein & O'Flaherty (1991) show the relationships among the concepts developed and familiar noncooperative concepts; Klein & O'Flaherty (1992) shows the crucial role of "imperfect policy formation" in Paretian time inconsistency; Hillier, Klein, and O'Flaherty (1992) show the crucial role of the prisoner's dilemma in Paretian commitment dominance.)

2. DEFINITIONS

The exposition of this section is organized in numbered subsections.

(2.1) An S-game is a four-tuple $\Sigma = (G, i, (F,C), s)$. (Item (F, C), "policy formation," is explained in Section 5 below.)

(2.2) The reference game G is an extensive form game.

(2.3) The ruler, i, and the public. Let I denote the index set of players of G, where $|I| = m+1$. Player i is called the ruler, and the set $I\backslash\{i\}$ of other players is called the public. For the remainder of this paper we let player 1 be the ruler (or i=1).

(2.4) Information assumption on G. Throughout this paper we make the assumption that a subgame originates at every ruler node. This is the common assumption in the time inconsistency literature.

(2.5) Plans. Let B' denote the set of player i's behavior strategies in G. In the S-game Σ, a subset B of B' is the set of permissible plans for the ruler, with generic element b. Section 5 on "policy formation" explains why the ruler may not have access to the complete set of behavior strategies, B'. Think of a plan as an announcement of what the ruler will do at each of her nodes.

(2.6) Public response function, s. Let S denote the set of behavior strategy m-tuples that exist in G for the public. Let s(b), the public response function,

denote the unique element of S that is picked out when plan b is announced. For the remainder of the paper we assume that for every plan b s(b) is a subgame perfect equilibrium in the m-player game induced by b.

(2.7) Ruler payoffs. U(b, d) denotes the ruler's payoff when she uses plan b and the public uses behavior strategy m-tuple d ε S. For U(b, s(b)) we will sometimes employ the summary payoff function u(b) := U(b, s(b)).

(2.8) Plan optimality. A plan b* is optimal iff for all b ε B, $u(b^*) \geq u(b)$.

(2.9) Local variations and subgames. For any ruler node v, let b_v denote the local strategy specified by b at v. Let $b \mid b'_v$ denote the ruler's behavior strategy that results if the local strategy assigned by b to node v is changed to b'_v while the local strategies assigned by b to other ruler nodes remain unchanged.

Let G(v) denote the subgame whose origin is ruler node v. Let b(v) denote the behavior strategy induced by plan b on G(v), and denote the ruler's payoff function on G(v) as $U_v(.)$, and her summary payoff function as $u_v(.)$. Let s(b(v)) denote the behavior strategy m-tuple induced on G(v) by s(b).

3. SEQUENTIAL IRRATIONALITY AND COMMITMENT DOMINANCE

In Figure 1, in the subgame G(v) the plan shown ([scratch 2's back]) is suboptimal. Similarly in Figure 2, in the subgame G(v) the plan shown

([break 2's back]) is suboptimal. We say that the plans of both examples are sequentially irrational because in each case choices are specified that are suboptimal in *some* subgame (not necessarily along the path of play!).

Formally, a plan b is sequentially rational iff for every ruler node v and every local strategy b'_v at v

$$u_v(\, b(v)\,) \geq u_v(\, b(v) \mid b'_v). \qquad (1)$$

Relation (1) says that what b specifies at each v is a best choice, where "best" is defined locally. To comprehend this definition think about applying (1) backward through the game. If a plan is not sequentially rational, it is sequentially irrational.

Using the ideas of sequential irrationality and plan optimality, we get a straightforward and natural standard for whether the ruler values commitment conveyance. Without commitment conveyance the ruler is restricted to sequentially rational plans. Therefore, were the ruler to have commitment conveyance and were all her optimal plans to be sequentially irrational, then we say she benefits from having commitment conveyance. We call this property "commitment dominance." Thus, when all optimal plans are sequentially irrational, the ruler faces commitment dominance. In both Figure 1 and Figure 2 the ruler faces commitment dominance.

Commitment Dominance and Schelling's Notion of Commitment

In *The Strategy of Conflict* (1960, 150), Schelling suggests that a commitment be thought of as a player's selective subtractions from her own

payoffs. In Figure 1 (restricting attention to the pure strategy optimal plan), the ruler conveys a commitment to [scratch 2's back]. Thus, following Schelling, we may say that she is convincing Player 2 that she has subtracted at least one unit of payoff from her own payoff at terminal node z, thereby making the plan credited. In Figure 2 , with optimal plan of [break 2's back], the ruler has convinced Player 2 that she has subtracted at least one unit of payoff from her own payoff at z.

Commitment dominance, then, means that a ruler would be positively benefited by having complete power to convey selective subtractions from her own payoffs. We embrace Schelling's idea of what it means to convey a commitment. Although our approach is fully consonant with Schelling's notion of selective subtraction, in our discussion there is no need to make further reference to it. Schelling's idea is implicit in optimal, sequentially irrational plans.

4. TIME INCONSISTENCY

We saw in Figure 2 that commitment dominance does not imply time inconsistency. The optimal plan [break 2's back] was sequentially irrational but it was also time consistent. Time consistency addresses desirable deviations from the original plan only along the path of play. Intuitively, a ruler faces time inconsistency iff along the path of an optimal plan she reaches a node where she would like to dissolve the original plan and revert to a subplan that is sequentially rational in that subgame. For play of a game in which reversion actually occurs (which implies that the original commitment was phony), there is a concatenation at the reversion point. The public is

startled at the reversion point.

Formally, the concatenated behavior strategy for the ruler $\kappa(b, v, b')$ is the behavior strategy that results from behavior strategy b if the behavior strategy induced by b on G(v), where v is a ruler node, is changed to b'(v) while the local strategies assigned by b to other ruler nodes remain unchanged.

Similarly, the concatenated behavior strategy m-tuple for the public $\gamma(d, v, d')$ is the behavior strategy m-tuple that results from behavior strategy m-tuple d if the behavior strategy m-tuple induced by d on G(v), where v is a ruler node, is changed to d'(v) while the local strategies assigned by d to other citizen nodes remain unchanged.

A plan b is time inconsistent iff there exists a ruler node v and some sequentially rational plan b' such that

$$U(\kappa(b, v, b'), \gamma(s(b), v, s(b'))) > u(b). \qquad (2)$$

A plan that is not time inconsistent is called time consistent.[4] Built into relation (2) is the along-the-path feature of time inconsistency: if node v is not along the original path of play, the reversion at v will not contribute to making the LHS of (2) greater than the RHS. Also built into the definition is the idea of the public being startled at v. Although they play according to s(b) only at citizens nodes that precede v, they play at those nodes under the belief that b will hold *for the entire game*.

Our definition of time inconsistency is a natural and faithful game-theoretic representation of what that term has always meant.

Tesfatsion, for example, is quite explicit about the along-the-path nature of time inconsistency. She says (1986, 25), an "economy is said to exhibit *inconsistency* if the competitive equilibrium path resulting from government reoptimization at some $t > 0$ is not a continuation of the competitive equilibrium path resulting from the initial government optimization at time 0." While Tesfatsion is careful to specify the path, most definitions of time inconsistency use temporal terms like "point in time," "later date," etc. Time periods do have meaning in pure game theory. Too often readers, and sometimes authors, have mistakenly interpreted "any point in time" as "any ruler node," and thus equated time inconsistency with the failure of subgame perfect equilibrium. This is an error, as has been pointed out by Fershtman (1990), McTaggart & Salant (1988), and Guiso & Terlizzese (1990). When definitions of time inconsistency refer to the initially optimal plan being no longer optimal at a future "point in time," this criterion applies only for the history that actually happens then, *not for any history that might have happened come that point in time.*

The only distinctive feature of our definition is that, following the interpretation of phony commitment conveyance, at a reached node the ruler can reannounce *only a sequentially rational subplan.* Alternatively one may wish to permit her to reannounce convincingly any subplan, and to fool the public repeatedly. The issue of the proper choice set at the point of deviation has scarcely arisen in the time inconsistency literature because in those models the government typically reaches a point of time inconsistency at its final decision node, so sequential rationality would be its preferred deviation even if it had a wider choice. Our decision to restrict reannouncements to

sequentially rational plans conforms to the proverb: "Fool me once, shame on you; fool me twice, shame on me." Once fooled, the public will not believe anything but a sequentially rational plan.[5] Previous definitions of time inconsistency have not addressed this issue, but it must be addressed whenever the game goes beyond the ruler's last move.

5. POLICY FORMATION

In the Kydland & Prescott (1977) Phillips curve model, a government commitment to inflation policy cannot take the form of any reaction function (with domain being the history of citizen employment decisions). In Fischer's saving-taxation model (1980), a government commitment to tax policy cannot take the form of any reaction function (with the domain being the history of citizen saving decisions). In Stackelberg duopoly, a leader commitment to output cannot take the form of any reaction function (with the domain being the follower output decision).

In the time inconsistency literature, the government is restricted to plans that take the form of a single magnitude that applies uniformly across all possible citizen histories standing at the government's "time to act." Hence the literature makes frequent use of the term "open loop" in describing the "rules," or commitment, regime. We need to account for the possible imperfection of policy formation.

What motivates imperfect policy formation? The familiar game theoretic reason why a player would have to take the same action at different nodes -- namely, imperfect information -- does not apply: if a government cannot

observe an individual citizen's savings, how can it tax them? The appropriate motivation is to realize that the real issue is not government action but government commitment conveyance, and there may be limits on the complexity of the commitments that the government can convey. Although anyone who has observed the hundred or so volumes of the U.S. Code Annotated may have doubts about this motivation, there is much intuition and a long tradition behind the assumption that government may face coarseness constraints in laying down policy.

To be precise, policy formation is described by a pair (F, C). F is a partition of the ruler's nodes into eligible subsets. We use the term eligible in Selten's sense (1975, 26): a set f of ruler nodes is eligible "if every play intersects [f] at most once, and if the number of alternatives at [v] is the same for every [v] ε [f]." We call these eligible subsets policy sets. Loosely speaking, a policy set is a set of nodes that the ruler must treat similarly when conveying a commitment. It may be useful to think of a policy set as a "point in time."

For any $f \varepsilon F$, let A_f be the set of all alternatives at the nodes in f. C is a partition of all the alternatives at all the ruler nodes; specifically, C partitions these alternatives into subsets c of all the A_f. Each of these subsets, c, must be eligible in Selten's sense (1975, 26): a subset c of A_f is eligible "if it contains exactly one alternative" at each node in the policy set.

The idea of policy formation is simpler than we are making it sound. A policy set f is a set of ruler nodes and a choice c is a choice at a policy set. The

idea of c is to define the choosing of the "same action" at every node in a policy set, as, for example, the taxing authority may be restricted to announcing a single tax rate that will prevail in the second period, no matter what history (or node) actually happens.

Now, a plan b for the ruler is eligible iff for every policy set f ε F either

(1) the set of alternatives b specifies is an element of C,

or (2) the plan is sequentially rational at every node in f.

("Sequential rationality at a node" means that relation (1) holds at that node.) Loosely speaking, programmers can think of option (1) as the "open-loop" option and option (2) as the "closed-loop" option.

B is the set of eligible plans, so policy formation restricts the set of behavior strategies of the reference game G that the ruler can convey a commitment to. We depict policy sets by connecting ruler nodes in a single policy set with a dashed line.

An example should clarify matters. Figure 4 portrays Stackelberg duopoly with three output levels available to each firm. (The outputs and payoffs shown correspond to Stackelberg and Cournot equilibria in a continuous variable model.[6]) Firm 1 is the leader, or ruler, and all of her nodes are contained in a single policy set. The three branches labelled [24], for example, form a single choice c at the policy set. She must announce such a c or declare sequential optimality at the policy set. Her best plan is to announce [24]. This

output level must be chosen at all three of the ruler's nodes because of the policy set. (Note that the plan is neither a pure promise nor a pure threat.) The plan elicits [12] in response. We can see that the Stackelberg equilibrium (shown by the arrows) is time inconsistent; the ruler would like to revert to [16].[7]

[Figure 4 here.]

If every policy set of an S-game is a singleton we say that the S-game satisfies perfect policy formation. Otherwise policy formation is imperfect.

A ruler would always prefer perfect to imperfect policy formation, because perfect policy formation expands the set of eligible plans. If the Stackelberg ruler in Figure 4 had perfect policy (blot out the dotted line), she could then announce a plan $b^* = [24, 24, 16]$ and enjoy payoff $u(b^*) = 640$. As the picture is drawn (imperfect policy formation) her optimal plan delivers only 576.

6. THE RULER'S WILLINGNESS TO PAY FOR COMMITMENT CONVEYANCE

When a ruler faces commitment dominant, how much is she willing to pay for the ability to convey a commitment? Suppose the price of commitment conveyance takes the form of a uniform deduction from every ruler payoff.[8]

To answer the question we must specify whether the commitment conveyance thus procurred conforms to Interpretation 1 or Interpretation 2 set out earlier. That is, we must specify whether commitment making is genuine or only apparent. Under Type 1 commitment conveyance, where a commitment to a plan must be genuine, the ruler is willing to pay up to and no more than

$$\delta_1 := u(b) - u(b'),$$

where b is an optimal plan and b' is a sequentially rational plan.

Under Type 2 commitment conveyance, where the commitment is only apparent and the ruler knows that she will have the opportunity to deviate, evaluating the ruler's willingness to pay is not so simple. Certainly δ_1 is a lower bound on the ruler's maximum willingness to pay, since the ruler can enjoy the payoff associated with actually sticking to an announced optimal plan. If an optimal plan b is time inconsistent, however, she would be willing to pay more than δ_1, because she can do even better than u(b) by deviating somewhere along the path. For example, in Figure 1, $\delta_1 = 3.25$ (= [0.25 * 5 + 0.75 * 4] - 1), but the ruler would be willing to pay up to 4 (= 5 - 1) for Type 2 commitment conveyance.

But now consider Figure 5. With Type 2 commitment conveyance, the ruler would not announce an "optimal" plan, like (L, l) or (L, r), but rather (R, r), *intending to deviate at k.* Thus we might say that the ruler is willing to pay 1 (= 3 - 2) for Type 2 commitment conveyance. We might suppose, however, that an announcement of (R, r) would arouse some suspicion. If player 2 has a shadow of a doubt about the authenticity of the ruler's commitment, he might conclude, in the spirit of the Intuitive Criterion of Cho & Kreps (1987) or forward induction (van Damme [1989]), that the ruler's commitment is phony, since a ruler limited to genuine commitments never would have a reason to annouce (R, r), whereas a Type 2 ruler might have a reason.

[Figure 5 here.]

Similarly, in the Stackelberg duopoly example treated in Section 5 and footnote 5, complete credulity on the part of the follower would cause a leader with Type 2 commitment conveyance to announce an output of 48, scaring the follower out of the market, and permitting the leader to revise to the pure monopoly output of 24. But again it is unreasonable to assume complete credulity when 48 is announced.

To formulate willingness to pay for Type 2 commitment conveyance, we assume that if the ruler announces a suboptimal plan the public disregards the announcement and proceeds as though it were common knowledge that the ruler is fixed on some specific sequentially rational plan b'. Thus, if the ruler announces a suboptimal plan the public "catches on" and the ruler receives the sequentially rational payoff. This restriction transforms s(:) into a new public response function, call it $\sigma(:)$. For any optimal plan b, $\sigma(b) := s(b)$; for any suboptimal b, $\sigma(b) := s(b')$.

Once the ruler announces any optimal plan b* she can take one of two tacks: she can exploit the most rewarding reversion opportunity along the path of (b*, s(b*)) , or she can stick to b*. The payoff from whichever tack is better, minus the sequentially rational payoff, is the ruler's willingness to pay for Type 2 commitment conveyance, denoted δ_2. Formally, let

$$\delta_{TI} := \max_{\substack{b^* \text{ is optimal} \\ v \in R \\ b' \text{ is sequentially rational}}} U\left(\kappa(b^*,v,b'), \gamma(\sigma(b^*), v, \sigma(b')) \right) - u(b')$$

The willingness to pay for Type 2 commitment conveyance is:

$$\delta_2 := \max[\delta_1, \delta_{TI}].$$

7. THE SAVINGS-TAXATION EXAMPLE

Here we give an S-game rendition of one of the familiar time inconsistency models: the saving-taxation model, exposited thoroughly by Fischer (1980) using a representative individual. In the first period citizens decide how much to save and in the second period they decide how much to work. In the first period the utilitarian government does nothing, and in the second period it levies taxes on labor income and accumulated savings to finance a public good.

In what the literature calls a "rules" regime, the government conveys a commitment to tax rates before the savings decision is made. The rules regime faces imperfect policy formation. The government can commit only to a tax plan that levies the same pair of tax rates at any "second period" node. That is, the government cannot make its tax plan contingent on citizen behavior. We are not arguing that this is a reasonable restriction; we are only describing the existing literature.

Alternatively, in what is called a "discretionary" regime, it is common knowledge that the government reoptimizes given the history of citizen decisions. In our language, the discretionary regime is a ruler restricted to sequentially rational plans.

Because savings taxation is nondistortionary the discretionary government will opt to tax savings heavily. Foreseeing this, citizens curtail savings in the first period. Everyone is better off under the rules regime because the government commits to moderate taxes on savings, inducing greater savings and hence more second-period consumption and public good provision. The externality driving the model is that each citizen does not consider the public-good benefit accruing to the other citizens when making his savings decision.

An S-game rendition is offered in Figure 6. There are two citizens (players 2 & 3) and a utilitarian government (player 1). Although Fischer's model uses continuous variables, we need only depict the choice levels that arise under the various regimes. The figure shows the two citizens simultaneously choosing between the Low savings induced by the discretionary regime and the High savings induced by the rules regime. The combinations of their choices imply four nodes for the government, which, in keeping with the literature, form a single policy set (shown by the dotted line).

[Figure 6 here.]

Each action at a government node represents a tax-rate pair for labor and accumulated savings. From left to right the four actions are:

t_{D1} - the sequentially rational tax-rate pair when each citizen has chosen Low savings.

t_{D2} - the sequentially rational tax-rate pair when one citizen has chosen Low savings and the other has chosen High savings.

t_{D3} - the sequentially rational tax-rate pair when both citizens have chosen High savings.

t_R - the optimal tax-rate pair when the government enjoys commitment conveyance.

(D subscripts are for "discretion," R for "rules.") Again, in Fischer's model the tax rates are continuous variables but we can simulate the continuous model by considering only the relevant choices. Note also that after tax-rates are confirmed individuals make a labor decision, which we do not depict because a unique equilibrium set of labor decisions will be implied by each history.

The discretionary regime cannot convey a commitment and, hence, is restricted to sequentially rational play, shown in Figure 6 by the double arrows. We assume that the citizens respond with a subgame perfect equilibrium in the game induced by the common knowledge of the government's plan. Hence the discretionary outcome is terminal node D. Notice that the citizen game induced by discretion is a prisoners' dilemma. This result is generalized by Hillier, Klein, and O'Flaherty (1992). There is always a prisoners' dilemma lurking behind Paretian commiment dominance.

A ruler with commitment conveyance will announce the tax-rate pair t_R, eliciting High savings from the citizens. Abiding by the plan would yield

outcome R, which Pareto dominates the discretionary outcome. The government would do even better if it could renege on its commitment, and, once at node z, both citizens would support such a reneging, which would yield the fooling outcome F.

Notice that if policy formation were perfect the ruler could announce an optimal plan that yielded the payoff of the time inconsistent reversion from the optimal plan shown in the figure (that is, payoff = 20). For example, she would announce (t_R, t_R, t_R, t_{D3}) and receive a payoff of 20. This plan would be *time consistent*. Klein & O'Flaherty (1992) generalize this result, proving that for any S-game with perfect policy formation some Paretian ruler has time consistent optimal plans. Thus imperfect policy formation lies at the heart of Paretian time inconsistency.

Figure 6 demonstrates the two points emphasized by the time inconsistency literature: (1) even a Paretian government may need commitment conveyance, and (2) there may be a conflict between optimality and consistency.

8. CONCLUSION

By designating a ruler who conveys a commitment prior to the play of a game, this paper offers a game-theoretic formulation of commitment and time consistency. We formulate the coarseness of policy formation, and discuss the ruler's willingness to pay for commitment conveyance. Our formulation is not proposed as a replacement for the current understanding of commitment and time consistency. Rather, we have tried to remain

faithful to the usages of Stackelberg analysis, Schelling's discussions, and the time inconsistency literature. We have tried to join various streams in a more rigorous, more general, and more complete framework.

S-game apparatus invites development and diverse application. Consider an S-game version of a one-shot sequential prisoner's dilemma with perfect policy formation. In pure plans, with the ruler moving second, the optimal plan induces mutual cooperation (as demonstrated by Thompson & Faith (1981, 372)). This is a true and familiar story. Perhaps daily we witness a situation in which a player moving second conveys a promise to return cooperation.

Notes

[1] Exceptions to the absence of commitment in pure theory include O'Flaherty's (1985) development of commitment as selective subtractions from one's own payoffs, and Thompson & Faith's (1980; 1981) study of commitment hierarchy. Simaan & Cruz (1973a, 1973b) offer a general discussion of Stackelberg play in non-zero sum games. Authors that incorporate commitment choices into a specific model include Lohmann (1990) and Flood & Isard (1989). An application of the commitment ideas developed here is Klein (1990).

[2] Schelling (1960, p177) pointed out: "randomization is evidently relevant to promises. If the only favors available to be promised are larger than necessary and are not divisible, a lottery that offers a specified probability of the favor's being granted can scale down the expected value of the promise and reduce the cost to the person making it."

[3] To be precise, when the genie appears the ruler gets to reannounce convincingly any sequentially rational plan. The new plan may differ from the original plan but still not specify a different local strategy where the genie appeared. See footnote 3 for further discussion.

[4] An important note of interpretation: Suppose b is time inconsistent at v and b_v is a mixed strategy. If the ruler reverts to an sequentially rational plan b' at v, and b'_v specifies an action which received positive probability under b_v, how would the citizens know that the ruler changed her plan? We must assume that the citizens are fully informed of the new plan; they see the new roulette wheel that is spun at v, as well as the wheels to be spun at successor ruler nodes. The authors are grateful to Stergios Skaperdas for pointing this out.

[5] None of the discussion of this paper depends on the decision to restrict b'(v) to sequential rationality. In Klein & O'Flaherty (1992) we use a weaker definition of time inconsistency which permits b'(v) to be any plan.

[6] Market clearing price = $100 - 2(y_1 + y_2)$. Marginal cost = 4. The Cournot

equilibrium is $y_1 = y_2 = 16$. The Stackelberg equilibrium is $y_1 = 24$, $y_2 = 12$.

[7]In the continuous space model the best deviation would be to 18. We could have included an action of 18 in the Figure but chose instead to minimize clutter.

[8] In this discussion we assume "SR-equivalence," which is a condition treating ties in ruler payoffs (Klein & O'Flaherty 1991). It is of small importance.

REFERENCES

Blanchard, Olivier Jean and Stanley Fischer [1989]: *Lectures on Macroeconomics* (Cambridge, MA: MIT Press).

Cho, In-Koo and David M. Kreps [1987]: "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179-221.

Fershtman, Chaim [1988]: "Fixed Rules and Decision Rules: Time Consistency and Subgame Perfection," *Economics Letters*, 191-194.

Fischer, Stanley [1980]: "Dynamic Inconsistency, Cooperation and the Benevolent Dissembling Government," *Journal of Economic Dynamics and Control*, 2, 93-107.

Flood, Robert P. and Peter Isard [1989]: "Monetary Policy Strategies," *IMF Staff Papers*, 612-632.

Guiso, Luigi and Daniele Terlizzese [1990]: "Time Consistency and Subgame Perfection: The Difference between Promises and Threats," Banca d'Italia Discussion Paper, no. 138.

Hillier, Brian, Daniel Klein, and Brendan O'Flaherty [1992]: "Policy Commitment and Welfare Gains," ms.

Hillier, Brian and James M. Malcomson [1984]: "Dynamic Inconsistency, Rational Expectations, and Optimal Government Policy," *Econometrica*, 52, 1437-1451.

Klein, Daniel B. [1990]: "The Microfoundations of Rules versus Discretion," *Constitutional Political Economy*, Fall, 1-19.

_____and Brendan O'Flaherty [1993]: "On the Game-Theoretic Rendering of Promises and Threats," *Journal of Economic Behavior and Organization*, 21, 295-314.

_____[1992]: "Contingent Plans, Time Consistency and Paretian Rulers," ms.

_____ [1991]: "Time Inconsistency and Related Concepts," ms.

Kydland, Finn and Edward Prescott [1977]: "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85, 473-493.

Lohmann, Susanne [1992]: "Optimal Commitment to Monetary Policy: Credibility Versus Flexibility," *American Economic Review*, 82, 273-286.

McTaggart, Douglas and David Salant [1988]: "Time Consistency and Subgame Perfect Equilibria in a Monetary Policy Game," ms.

Nalebuff, Barry and Martin Shubik [1988]: "Revenge and Rational Play," Princeton University discussion paper 138.

O'Flaherty, Brendan [1985]: *Rational Commitment: A Foundation for Macroeconomics* (Durham, NC: Duke University Press).

Rasmusen, Eric [1989]: *Games and Information* (Cambridge, MA: Basil Blackwell).

Schelling, Thomas C. [1960]: *The Strategy of Conflict* (Cambridge: Harvard University Press).

Selten, Reinhart [1975]: "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25-55.

Simaan, M. and J. B. Cruz, Jr. [1973a]: "On the Stackelberg Strategy in Nonzero-Sum Games," *Journal of Optimization Theory and Applications*, 11, 533-555.

_____ [1973b]: "Additional Aspects of the Stackelberg Strategy in Nonzero-Sum Games," *Journal of Optimization Theory and Applications*, 11, 613-626.

Tesfatsion, Leigh [1986]: "Time Inconsistency of Benevolent Government Economies," *Journal of Public Economics*, 25-52.

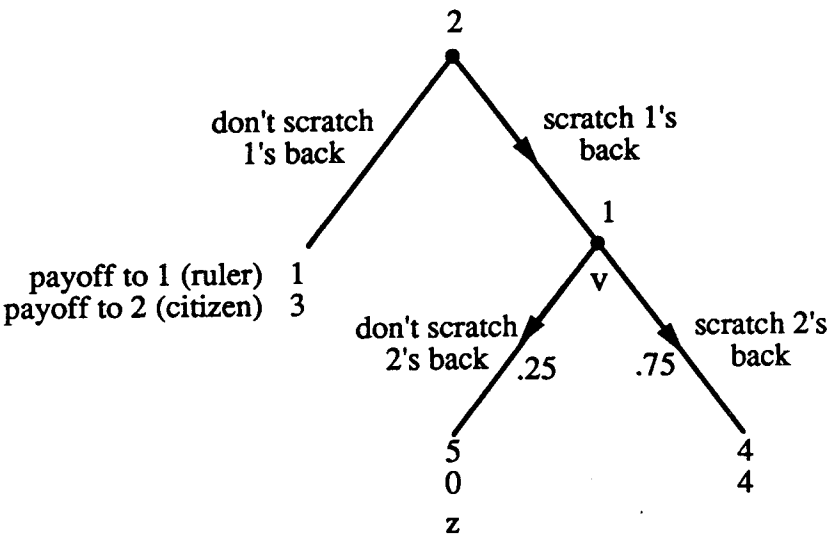
Thompson, Earl A. and Roger L. Faith [1980]: "Social Interaction under Truly Perfect Information," *Journal of Mathematical Sociology*, 7, 181-197.

_____ [1981]: "A Pure Theory of Strategic Behavior and Social Institutions," *American Economic Review*, 71, 366-380.

van Damme, Eric [1989]: "Stable Equilibria and Forward Induction," *Journal of Economic Theory*, 48, 476-496.

Figure 1

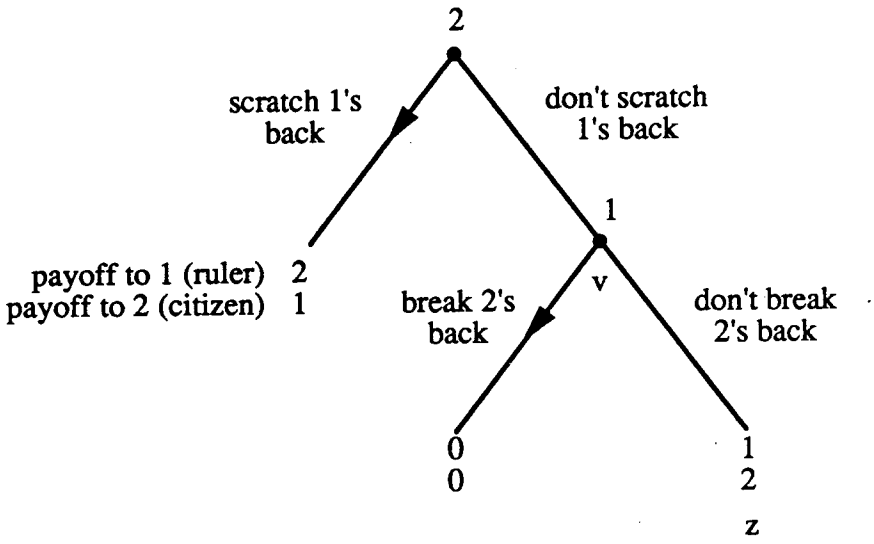
"If you scratch my back I'll scratch your back."



b and s(b) are shown by the arrows; b specifies (don't scratch) with .25 probability and (scratch) with .75 probability. b is optimal and sequentially irrational, and it is time inconsistent because at v 1 would like to switch to sequentially rational play (don't scratch).

Figure 2

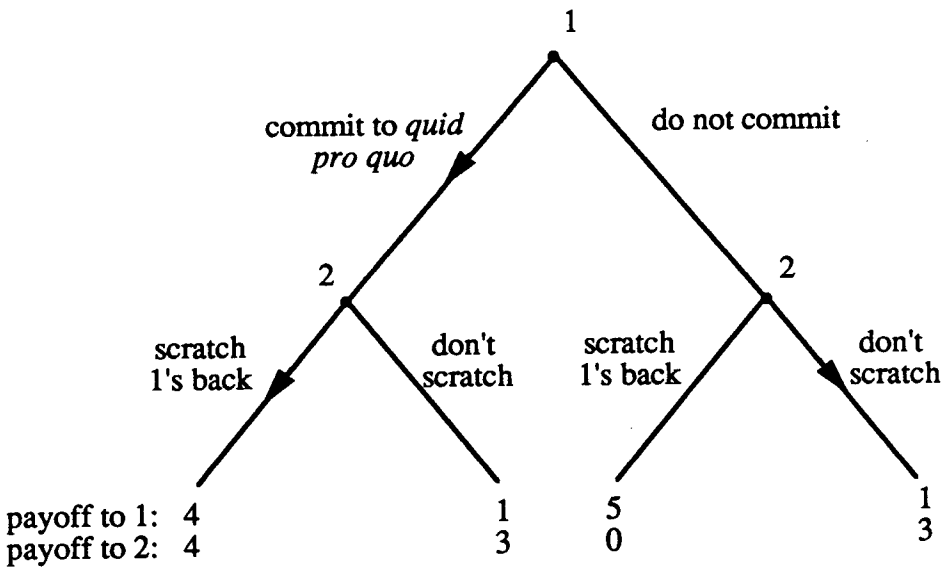
"Scratch my back or else I'll break your back."



b and s(b) are shown by the arrows. b is optimal, sequentially irrational, and time consistent.

Figure 3

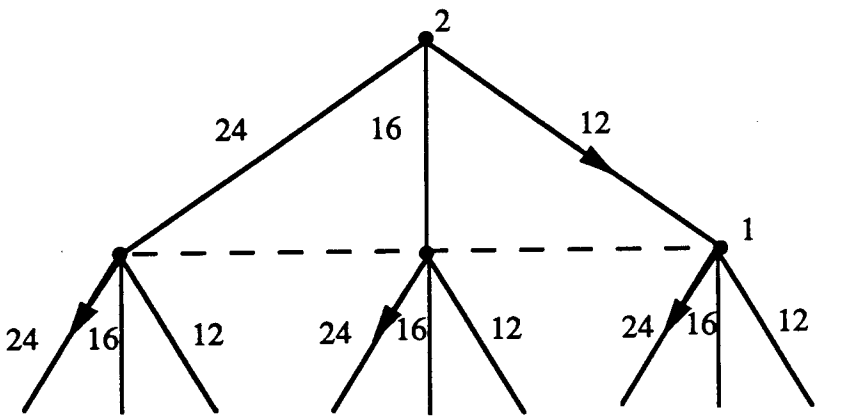
A standard game formulation of
the commitment story of Figure 1.



The arrows show the unique perfect equilibrium.

Figure 4

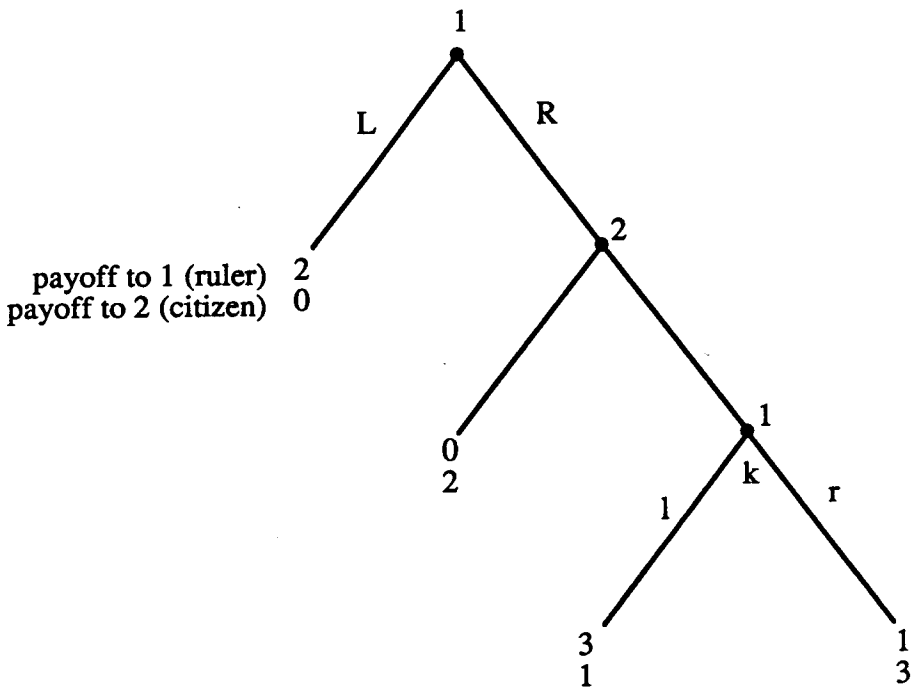
Stackelberg Duopoly.



payoff to 1 (ruler)	0	256	288	384	512	480	576	640	576
payoff to 2 (citizen)	0	384	576	256	512	640	288	480	576

Every ruler node is contained in a single policy set (shown by the dotted line). b and s(b) are shown by the arrows. b is optimal and time inconsistent.

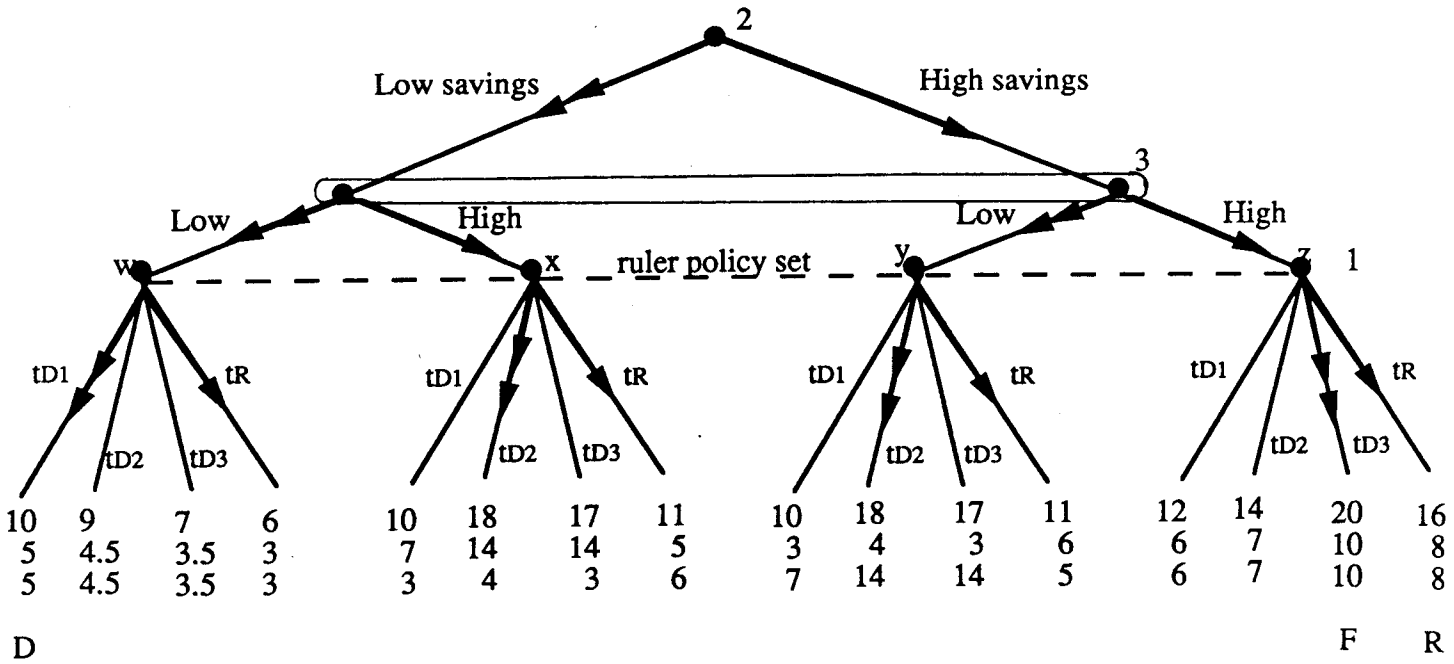
Figure 5



If you were Player 2, how would you respond to the ruler's announcemnt of (R, r)?

Figure 6

The savings-taxation model with two citizens.



Payoffs are listed in the order: player 1 (ruler)
player 2
player 3

Double arrows show strategies under sequential rationality (or "discretion"). Notice that under this regime the citizens are playing a prisoner's dilemma.

Single arrows show strategies under the optimal plan ("rules" regime). The optimal plan is time inconsistent.

Notice that if the ruler had perfect policy formation (blot out the dashed line), she could announce tR at w, x, and y and tD3 at z. That plan would be time consistent and arrive at node F.