Errata for “The Turnover Trap”*

Scott Wolford
Emory University
mwolfor@emory.edu

November 28, 2007

The proofs for “The Turnover Trap: New Leaders, Reputation, and International Conflict” (Wolford, 2007) omit some parameter restrictions that support the uniqueness of the presented Perfect Bayesian Equilibrium. Here, I clarify (1) the character of the parameter space that supports the equilibrium and (2) the extent to which the presented equilibrium is unique given those parameter values. I also rectify a typographical error in the proofs.

Interior proposals

First, the derived equilibrium proposals $x_{t+1}^r$ and $x_t$ are assumed to be interior—i.e., $0 < \{x_{t+1}^r, x_t\} < 1$—which requires some restrictions on the relative resolve of leaders $A_k$ and $B_k$. I demonstrate this in terms of $A_k$’s resolve, $a > 0$, ensuring that he faces a true risk-return tradeoff, or that $A_k$ is not so irresolute that he is unwilling to risk war with $B_k$ and and not so resolute that he will always pursue war.

First, since $x_{t+1}^S = x_{t+1}^d$ (as stated on p. 784), each of these proposals is interior when $2p + \bar{b} - 2 < a < 2p + \bar{b}$. Second, $x_{t+1}^h$ is interior when $2p + b_t - 2 < a < 2p - b_t$. Since $\bar{b} > b_t$, Inequality (1) guarantees that all three proposals at time $t + 1$ are interior:

$$2p + \bar{b} - 2 < a < 2p + b_t. \quad (1)$$

Since $b_t$, defined as a function of $x_t$ on p. 785, is an endogenous variable, note that it increases as a function of $B_k$’s incentives to bluff (the probabilities of both surviving and winning a war) and decreases as a function of incentives to respond honestly (the probability of surviving a peaceful settlement).

The interiority of proposals at time $t + 1$ is not strictly necessary to ensure that the “screening” character of the PBE presented in the text is maintained: if, for example, values of $a$ outside those defined by Inequality (1) support corner solutions $x_{t+1}^d = 1$ and/or $x_{t+1}^h = 0$, it is still true that $U_{B_1}(A_k^d) > U_{B_1}(A_k^h)$. I simply impose this restriction in order to focus the presentation on a single equilibrium rather than other, generally similar, equilibria.

* Sincere thanks to Alexandre Debs for identifying this issue in the published version of the article.
I establish the interiority of \( x_t \) with bounds on \( A_1 \)'s resolve, but given the cumbersome nature of the values that satisfy \( 0 < x_t < 1 \), I present them in more general form. First, Equation (A.4) shows that \( x_t \) is linear in \( a \), which can also be verified by confirming that the first derivative with respect to \( a \) is constant. This implies the existence of two values of \( a \) such that \( 0 < x_t < 1 \) when \( a < x_t < a \).

Given these values and the constraints in Inequality 1, we can state the parameters supporting the interiority of all proposals as

\[
\max \{2p + b - 2, a\} < a < \min \{2p + b_t, a\}.
\]

**Out-of-equilibrium beliefs and uniqueness**

Since all actions are taken with positive probability on the equilibrium path in the presented PBE, no out-of-equilibrium beliefs need be specified to prove existence. Uniqueness, however, requires examining the sustainability of potential alternative equilibria, which does require a treatment of out-of-equilibrium beliefs.

On p. 785, the proofs reject an alternative pooling equilibrium in which all types of \( B_1 \) reject \( x_t \) and go on to receive a proposal \( x_S^{t+1} \) consistent with \( A_k \)'s inability to update its beliefs. As an alternative to rejection, some irresolute types can reveal themselves and receive \( x_d^{t+1} \) in the semi-separating PBE. Therefore, this pooling equilibrium is unsustainable—and players do better in the presented PBE—when (a) \( A_k \) believes that it faces a sufficiently irresolute type given an unexpected acceptance, or \( b \in [b_t, b] \), (b) \( A_1 \) sets an initial offer \( x_t \), as identified in Equation (A.4), that encourages acceptance by weak types, and (c) proposals at time \( t+1 \) are interior as defined by Inequality (1).

The question is whether other out-of-equilibrium beliefs \( A_k \) might hold would support other pooling PBEs that overlap the semiseparating equilibrium and render it non-unique. Consider an equilibrium in which all types of \( B_1 \) reject \( x_t \), after which, by Inequality (1), \( A_k \) proposes \( x_S^{t+1} \). Such an equilibrium must be sustained by \( A_k \)'s behavior in the event that it observes an unexpected acceptance, which is a function of its out-of-equilibrium beliefs.

To establish some plausible beliefs for \( A_k \), assume that, in the case of acceptance, its beliefs support a continuous range of types that includes the most irresolute type \( \bar{b} \).¹ For rejection to be a best reply for all types of \( B_1 \), it must be the case that there is no incentive for irresolute types to identify themselves, which can be shown in two steps. First, as long as \( A_k \) believes that at least the most irresolute type \( \bar{b} \) is among the types that would accept, the proposal at time \( t+1 \) will be identical to \( x_S^{t+1} \) and \( x_d^{t+1} \), since proposals in the second crisis are conditioned on the most irresolute type \( A_k \) believes that it faces (see p. 784). Second, given that outcomes at time \( t+1 \) are the same for acceptors and rejectors, a type will reject so long as war is not prohibitively costly; thus, there will exist some types \( b \geq b_c \), where \( b_c \) is a function of relative survival probabilities and the military balance,

¹I reject as implausible out-of-equilibrium beliefs in which \( A_k \) believes that more resolute types, say those including and approaching \( \bar{b} \), would accept while irresolute types like \( \bar{b} \) do not. Another potential case for out-of-equilibrium beliefs would allow for \( A_k \) to believe that only the most irresolute type, \( \bar{b} \) would accept, leading \( A_k \) to propose \( x_{t+1} = p + \bar{b} \) in the second crisis. All types but \( \bar{b} \) would reject this proposal, which should rule defection out even for \( \bar{b} \) since \( x_{t+1}^{t+1} < p + \bar{b} \).
that will accept a proposal of $1 - x_t = 0$ rather than pay the costs of fighting in the present. Since such types always accept, assume for the rest of the discussion that $b < b_c$.

Given that $B_1$’s behavior is a best reply, is $x_t = 1$ an optimal proposal for $A_1$? His expected utility is given by

$$U_{A_1}(x_t = 1) = \int_{b}^{b_{t+1}} (p[1-a + \alpha_W (\beta_L(p-a) + (1-\beta_L)U_{A_1}(B_2))] + (1-p)[-a + \alpha_L (\beta_L(p-a) + (1-\beta_L)U_{A_1}(B_2))] \theta db + \int_{b_{t+1}}^{b} (p[1-a + \alpha_W (\beta_L x_{t+1}^S + (1-\beta_L)U_{A_1}(B_2))] + (1-p)[-a + \alpha_L (\beta_L x_{t+1}^S + (1-\beta_L)U_{A_1}(B_2))] \theta db, \quad (2)$$

which includes guaranteed war in the first crisis and, if $B_1$ survives, a probability distribution defined by his priors, $\theta_{t+1}^b = \theta_t^b$, over war and the acceptance of $x_{t+1}^S$.

Since $A_1$ can set his own proposal for $x_t$, it is necessary to compare $U_{A_1}(x_t = 1)$ to some interior proposal $x_t \in (0, 1)$. Given an interior proposal, with which $A_1$ can induce some irresolute types of $B_1$ to accept, we know that $A_1$’s optimum is $x_t$ as defined by Equation (A.4). Therefore, $U_{A_1}(x_t)$ as defined by Equation (A.3) provides the relevant comparison to Equation (2).

Begin with outcomes at $t+1$. First, since $b_{t+1}^d = b_{t+1}^S \iff x_{t+1}^d = x_{t+1}^S$ and $x_{t+1}^h > p - a \forall b_t > 0$, $A_1$ always does better in the second crisis in the semiseparating equilibrium than in the pooling equilibrium. Second, since $x_t > p - a$, $A_1$ has an incentive to induce acceptance rather than rejection as long as the probability of survival following military victory is not too high. In other words, there exists some value $\bar{\alpha}_W$ above which $A_1$ can increase his probability of survival following a war to such an extent that even suboptimal distributive outcomes in the first and second crises are acceptable. This value, which is cumbersome to present, is quite high and available from the author upon request. Therefore, when $\alpha > \bar{\alpha}_W$, $A_1$ prefers to set $x_t = 1$ and induce a pooling equilibrium. However, for all other values $\alpha_W \leq \bar{\alpha}_W$ and for plausible out-of-equilibrium beliefs, $A_1$ prefers instead to seek out a semiseparating equilibrium as presented in “The Turnover Trap.”

**Typographical error**

Finally, p. 786 contains a typographical error when stating the equilibrium probability of conflict for comparative statics analysis, which should be written as

$$Pr(\text{Conflict}) = \frac{b_t - b}{b - b_t}. \quad (3)$$

**References**