# Supplementary Appendix to: When is Reputation Bad?

Jeffrey Ely
Drew Fudenberg
David K. Levine[1]

November 22, 2005

In Ely, Fudenberg and Levine [2005], hereafter EFL, we defined an action to vulnerable to a temptation if conditional on participation by the short-run players, the temptation lowers the probability of all "bad" signals, and increases the probability of all other signals. EFL's first bad reputation result requires an exit minmax condition. If the temptation satisfies the stronger property that the relative probability of the other signals remains constant, then the assumption of exit minmax can be weakened. We define a strong temptation:

**Definition 3S:** *An action* $a$ *is* vulnerable to a strong temptation relative to a set of signals $\widehat{Y}$ *if there exists a number* $\underline{\rho} > 0$ *and an action* $d$ *such that*

*1) If* $b \notin E$, $\hat{y} \in \widehat{Y}$ *then* $\rho(\hat{y} \mid d,b) \leq \rho(\hat{y} \mid a,b) - \underline{\rho}$

*2) If* $b \notin E$ *and* $y, y' \notin \widehat{Y} \cup Y^E$ *then* $\dfrac{\rho(y \mid d,b)}{\rho(y' \mid d,b)} = \dfrac{\rho(y \mid a,b)}{\rho(y' \mid a,b)}$.

*3) For all* $e \in E$, $u^L(d,e) \geq u^L(a,e)$.

*The action* $d$ *is called a* strong temptation.

This condition lets us prove an analog of lemma 3 in EFL which we prove here.

[1] Departments of Economics, Northwestern University, Harvard University and UCLA.

**Lemma 3S:** *In a participation game, if* $\beta(h_t)\{E\} = 1$ *or* $\beta(h_t)\{E\} < 1$ *and* $\alpha_0(h_t)(f) > 0$ *for some friendly action* $f$ *that is vulnerable to a strong temptation of size* $\underline{\rho}$*, then*

$$v(h_t) \leq \max_{y \in Y(h_t)} (1-\delta)\overline{u}\ (y, \tilde{\rho}) + \overline{\delta}\,(y, \tilde{\rho}) v(h_t, y)$$

where

$$\overline{u}(y,\rho) = \begin{cases} \left(1 + \dfrac{1}{\left|\widehat{Y}\right| \tilde{\rho}}\right) U^L & \text{if } y \in \widehat{Y} \\ \hat{u}^L & \text{otherwise} \end{cases}$$

$$\text{and } \overline{\delta}\,(y,\rho) = \begin{cases} \delta\left(1 + \dfrac{1}{\left|\widehat{Y}\right| \tilde{\rho}}\right) & y \in \widehat{Y} \\ \delta & \text{otherwise} \end{cases}$$

*Proof:* The derivation of equation (1.5) is unchanged

$$\begin{aligned} &(1-\delta)U^L + \delta \textstyle\sum_{\widehat{y} \in \widehat{Y}} \big[\rho(\widehat{y} \mid f, \beta_{-E}) - \rho(\widehat{y} \mid d, \beta)\big] v(h_t, \widehat{y}) \\ &\geq \delta \Big[\textstyle\sum_{y \in Y / \widehat{Y}} \big[\rho(y \mid d, \beta_{-E}) - \rho(y \mid f, \beta_{-E})\big] v(h_t, y)\Big] \end{aligned} \tag{1.5}$$

Since the good signals are changed proportionately by the temptation, it follows that

$$\rho(y \mid b, \beta_{-E}) = \frac{\rho(Y \setminus \widehat{Y} \mid d, \beta_{-E})}{\rho(Y \setminus \widehat{Y} \mid f, \beta_{-E})} \rho(y \mid f, \beta_{-E})$$

for each $y \in Y \setminus \widehat{Y}$. Thus,

$$\begin{aligned} &\textstyle\sum_{y \in Y \setminus \widehat{Y}} \big[\rho(y \mid d, \beta_{-E}) - \rho(y \mid f, \beta_{-E})\big] v(h_t, y) \\ &= \textstyle\sum_{y \in Y \setminus \widehat{Y}} \left[\left(\dfrac{\rho(Y \setminus \widehat{Y} \mid b, \beta_{-E})}{\rho(Y \setminus \widehat{Y} \mid f, \beta_{-E})} - 1\right) \rho(y \mid f, \beta_{-E})\right] v(h_t, y) \\ &= \left(\frac{\rho(Y \setminus \widehat{Y} \mid d, \beta_{-E})}{\rho(Y \setminus \widehat{Y} \mid f, \beta_{-E})} - 1\right) \textstyle\sum_{y \in Y \setminus \widehat{Y}} \rho(y \mid f, \beta_{-E}) v(h_t, y) \\ &= \left(\frac{\rho(Y \setminus \widehat{Y} \mid d, \beta_{-E}) - \rho(Y \setminus \widehat{Y} \mid f, \beta_{-E})}{\rho(Y \setminus \widehat{Y} \mid f, \beta_{-E})}\right) v(h_t, Y \setminus \widehat{Y}) \end{aligned}$$

where $v(h_t, Y \setminus \widehat{Y})$ is the expected continuation value after playing $f$ and observing a signal in $Y \setminus \widehat{Y}$. Substituting into (1.5),

$$
\begin{aligned}
&(1-\delta)U^L + \delta\textstyle\sum_{\widehat{y}\in\widehat{Y}}\big[\rho(\widehat{y}\mid f,\beta_{-E}) - \rho(\widehat{y}\mid d,\beta_{-E})\big]v(h_t,\widehat{y}) \\
&\geq \delta\left[\frac{\rho(Y\setminus\widehat{Y}\mid d,\beta_{-E}) - \rho(Y\setminus\widehat{Y}\mid f,\beta_{-E})}{\rho(Y\setminus\widehat{Y}\mid f,\beta_{-E})}\right]v(h_t,Y\setminus\widehat{Y}) \\
&\geq \delta\big[\rho(Y\setminus\widehat{Y}\mid d,\beta_{-E}) - \rho(Y\setminus\widehat{Y}\mid f,\beta_{-E})\big]v(h_t,Y\setminus\widehat{Y})
\end{aligned}
$$

Set

$$v(h_t,\widehat{Y}) = \max_{y\in\widehat{Y}} v(h_t,y)\,.$$

From the fact that *d* reduces the probability of every bad signal by a positive amount,

$$
\begin{aligned}
&\big[\rho(\widehat{y}\mid f,\beta_{-E}) - \rho(\widehat{y}\mid d,\beta_{-E})\big]v(h_t,\widehat{y}) \\
&\leq \big[\rho(\widehat{Y}\mid f,\beta_{-E}) - \rho(\widehat{Y}\mid d,\beta_{-E})\big]v(h_t,\widehat{Y})
\end{aligned}
$$

for each $\widehat{y}\in\widehat{Y}$. Thus,

$$
\begin{aligned}
&(1-\delta)U^L + \delta\big[\rho(\widehat{Y}\mid f,\beta_{-E}) - \rho(\widehat{Y}\mid d,\beta_{-E})\big]v(h_t,\widehat{Y}) \\
&\geq \delta\big[\rho(Y\setminus\widehat{Y}\mid d,\beta_{-E}) - \rho(Y\setminus\widehat{Y}\mid f,\beta_{-E})\big]v(h_t,Y\setminus\widehat{Y})
\end{aligned}
\tag{1.6}
$$

which implies

$$
\begin{aligned}
v(h_t,Y\setminus\widehat{Y}) &\leq \frac{(1-\delta)U^L + \delta\big[\rho(\widehat{Y}\mid f,\beta_{-E}) - \rho(\widehat{Y}\mid d,\beta_{-E})\big]v(h_t,\widehat{Y})}{\delta\big[\rho(Y\setminus\widehat{Y}\mid b,\beta_{-E}) - \rho(Y\setminus\widehat{Y}\mid f,\beta_{-E})\big]} \\
&\leq \frac{(1-\delta)U^L + \delta v(h_t,\widehat{Y})}{\delta\big[\rho(Y\setminus\widehat{Y}\mid d,\beta_{-E}) - \rho(Y\setminus\widehat{Y}\mid f,\beta_{-E})\big]}
\end{aligned}
$$

where the second line uses the fact that $v(h_t,\widehat{Y}) \geq 0$.

Because $d$ lowers the probability of all bad signals by at least $\underline{\rho}$, it raises the total probability of the remaining signals by at least $\big|\widehat{Y}\big|\underline{\rho}$, that is, $\big[\rho(Y\setminus\widehat{Y}\mid d,\beta_{-E}) - \rho(\widehat{Y}\setminus Y\mid f,\beta_{-E})\big] \geq \big|\widehat{Y}\big|\underline{\rho}$. This and the fact that the numerator on the right hand side of the previous inequality is non-negative gives

$$v(Y\setminus\widehat{Y}) \leq \frac{(1-\delta)U^L + \delta v(h_t,\widehat{Y})}{\delta\big|\widehat{Y}\big|\underline{\rho}}$$

Finally, we conclude:

$$
\begin{aligned}
&(1-\delta)u^L(f,\beta_{-E}) + \delta\Big(\textstyle\sum_{\widehat{y}\in\widehat{Y}} \rho(\widehat{y} \mid f,\beta_{-E})v(h_t,\widehat{y}) + \rho(Y\setminus\widehat{Y} \mid f,\beta_{-E})v(h_t,Y\setminus\widehat{Y})\Big)\\
&\leq (1-\delta)u^L(f,\beta_{-E}) + \delta\Big(\rho(\widehat{Y} \mid f,\beta_{-E})v(h_t,\widehat{Y}) + \rho(Y\setminus\widehat{Y} \mid f,\beta_{-E})v(h_t,Y\setminus\widehat{Y})\Big)\\
&\leq (1-\delta)u^L(f,\beta_{-E}) + \delta\Big(v(h_t,\widehat{Y}) + v(h_t,Y\setminus\widehat{Y})\Big)\\
&\leq (1-\delta)U^L + \delta v(h_t,\widehat{Y}) + \delta\left(\frac{(1-\delta)U^L + \delta v(h_t,\widehat{Y})}{\delta\left|\widehat{Y}\right|\underline{\rho}}\right)\\
&= (1-\delta)\left(1+\frac{1}{\left|\widehat{Y}\right|\underline{\rho}}\right)U^L + \delta\left(1+\frac{1}{\left|\widehat{Y}\right|\underline{\rho}}\right)v(h_t,\widehat{Y})\\
&= \max_{y\in\widehat{Y}}(1-\delta)\overline{u}(y,\underline{\rho}) + \overline{\delta}(y,\underline{\rho})v(h_t,y)
\end{aligned}
$$

The conclusion of the proof is now identical to that of Lemma 3: if $\lambda = 0$ then $v(h_t) = v(h_t,f,\beta_{-E})$ and if $\lambda \in (0,1)$ then $v(h_t) \leq \max\left\{v(h_t,f,\beta_E), v(h_t,f,\beta_{-E})\right\}$.

☑