The Reputation Trap

David K. Levine
Easy to Lose and Hard to Gain

“Glass, china, and reputation are easily cracked, and never well mended.” Benjamin Franklin[?]

rationale:

• with a good reputation people will be eager to do business so if they are cheated it will quickly become known

• nobody wants to do business with a bad reputation so even if honest behavior takes place nobody will find out
adverse event -> loss of reputation
  • difficulty of restoration provides little incentive for honesty
  • bad reputation will deservedly remain so long after the circumstances that caused it are gone
circumstances dictate honesty -> long time before others find out
  • once they do reputation is restored
  • even after the circumstances dictating honesty are gone it will be desirable to continue to be honest to avoid losing reputation
two otherwise identical individuals may find themselves with entirely different incentives for honesty because of an adverse or beneficial event that happened in the distant past
An Application: Persistence of Dysfunction

political economy of culture and institutions: the persistence of dysfunctional cultures

- Acemoglu-Robinson: four centuries
- Bignoni et al: nine centuries
- Dell-Querubin: century and a half (strong evidence)
Rapid Change

side by side with the survival of dysfunctional norms we see abrupt change over periods of a few decades

• religion: 1990s Ireland from Catholic to secular
• language: 1900-1915 number of Hebrew speaker grew from a few hundred to more than 30,000
• economic culture: could Nigerians enrich themselves by replacing their own culture with Japanese culture

how did Japanese culture came to be what it is?

• 1868 Japan: culture, technology, and standard of living similar to medieval Europe
• 1904: transformed into a modern industrial state

40 years of the Meiji era affected virtually every aspect of life in Japan
Reputation Trap?

do good social norms for economic success revolve around good treatment of immigrants and foreign investors?

• poor treatment of foreigners won’t get much immigration or foreign investment

• so no urban centers of production and innovation

even if treatment of outsiders is improved nobody is likely to find out could be a reputation trap

given the enormous disparity in income between Nigeria and Japan is it only adjustment cost?
The Model

overlapping generations of finitely lived players; two player roles

- player 1 long-run who lives many periods
- player 2 a mass of short-run players who live a single period

stage game

**long-run player:** make a costly investment?

\[ a_1 \in \{0, 1\} \text{ with } 1 \text{ meaning to invest and the cost being } ca_1 \text{ where } 0 < c < 1 \]

**short-run player** without observing the investment choice:

enter \( a_2 = 1 \) or

stay out \( a_2 = 0 \)
Short-run Player Payoffs

• $0$ for staying out
• $-1$ for entering when no investment has been made
• $V > 0$ for entering when the investment has been made
Long-Run Player Types

three types $\tau \in \{b, n, g\}$

- $g$ means “good” (a beneficial event)
- $b$ means “bad” (an adverse event)
- $n$ means “normal”

player type is fixed during the lifetime of the player.
players care only about expected average utility during their lifetime

good and bad types are behavioral types
  - good type always invests
  - bad type never invests

payoff of the normal type is $a_2 - ca_1$
Life of the Long Run Player

Life of a long-run player is stochastic

- with probability $\delta$ the player continues for another period
- with probability $1 - \delta$ is replaced.

When a long-run player is replaced the type may change

Probability type $\tau$ is replace by a type $\sigma \neq \tau$ is

$$Q_{\tau \sigma} \epsilon / (1 - \delta) > 0$$

Interested in the case in which types are persistent – that is, in which $\epsilon$ is small
beginning of each period a public signal $z$ of the previous period is observed

- takes on one of three values: 1, 0, $\mathcal{N}$

If entry took place last period

- signal is equal to last-period long-run player investment

If the short-run player stayed out last period

- probability $1 \geq \pi > 0$ the signal is equal to last period long-run player investment

- probability $1 - \pi$ the signal is $\mathcal{N}$

when the short-run player stays out less information generated about the behavior of the long-run player
Strategies

game begins with an initial draw of the public signal $z(1)$ and private type $\tau(1)$ from the common knowledge distribution $\mu_{z,\tau}(1)$

players are only aware of events that occur during their lifetime

strategy of normal type of long-run player: investment probability $\alpha_1(h, t, T)$ function of private history $h$, calendar time $t$, and generation $T$

strategy for the short-run player: probability of entering $\alpha_1(z, t)$ as a function of the beginning of period signal and calendar time $t$

study Nash equilibria

generic assumption on $c$
Short-run Player Beliefs and Time Invariant Equilibrium

- short-run players stay out and normal type does not invest
- eventually it becomes likely that the long-run player has migrated back to a “normal” type
- short-run players and long-run player coordinate
- “today is the day”
- short-run enters and normal long-run invests
- good reputation restored

This type of coordination is not terribly plausible: requires that both players agree about the exact timing of events in the long-distant past.

Rule this out by assuming that equilibrium short-run player strategies and beliefs are independent of calendar time.

Implicit in Markov equilibrium, but here the long-run player strategies are not constrained.
equilibrium strategy for a short-run player now a time invariant probability of entering $\alpha_2(z) \in [0, 1]$

long-run player faces a well-posed Markov decision problem

$P(1|z, 1) = P(0|z, 0) = \alpha_2(z) + (1 - \alpha_2(z))\pi$ and

$P(N|z, a_1) = (1 - \alpha_2(z))(1 - \pi)$

$V(\alpha_2) = \max_{a_1}(1 - \delta) [\alpha_2 - ca_1] + \delta \sum_{z'} P(z'|z, a_1)V(\alpha_2(z'))$

set of best responses determined by $\alpha_2(z)$

best response $\alpha_1(y_t, t, T_t)$ must lie in this set

hence time invariant beliefs of the short run player $\alpha_1(z)$ must also be a best response.
Short-Run Player Optimization

beliefs $\mu_{z_T}(t)$
observe $z_t$
find $\mu_{T|z_t}(t)$

combine with $\alpha_1(z_t)$ to find $\mu^1(z_t, t)$: beliefs about probability of long-run player investment

$\alpha_2(z_t)$ must be a best response to $\mu^1(z_t, t)$
Evolution of Beliefs

\[ \mu_{zT}(t) \] determined by

initial condition \[ \mu_{zT}(1) \]

beliefs of the short-run player about the probabilities with which earlier normal-type long-run and short-run players chose actions \( \alpha_1(z) \), \( \alpha_2(z) \)

does not depend on the actual choice of those actions or the earlier signals as those are not observed

\( \bar{\mu}(t) \) vector of \( \mu_{zT}(t) \)

\[ \bar{\mu}(t + 1) = A \bar{\mu}(t) \]

\( A \) Markov transition matrix coefficients determined by \( \alpha_1(z) \), \( \alpha_2(z) \) and \( \pi, Q, \epsilon \)
Time Invariant Beliefs

time invariant beliefs : \( \overrightarrow{\mu}(t + 1) = \overrightarrow{\mu}(t) \)

if and only if the initial condition \( \mu_{zt}(1) \) a stationary distribution of \( A \).

cannot have arbitrary initial short-run player beliefs \( \mu_{zt}(1) \), only initial beliefs that are consistent with the strategies of the players and the passage of time.
Time Invariant Equilibrium

a triple $(\alpha_1(z), \alpha_2(z), \mu_{zt})$

$\alpha_1(z)$ solves the Markov decision problem induced by the short-run player strategy $\alpha_2(z)$

$\mu_{zt}$ a stationary distribution of the matrix $A$ determined by $\alpha_1(z), \alpha_2(z), Q, \epsilon$

$\alpha_2(z)$ is a best response to beliefs about long-run player action $\mu^1(z)$ determined from $\alpha_1(z), \mu_{zt}$. 
Equivalence

$z(y)$ the most recently observed signal by the long-run player in the history $y$

**Theorem:** If $(\alpha_1(z), \alpha_2(z), \mu_{z\tau})$ is a time invariant equilibrium then the strategies $\alpha_1(y, t, T) = \alpha_1(z(y)), \alpha_2(z, t) = \alpha_2(z)$ are a Nash equilibrium with respect to the initial condition $\mu_{z\tau}(1) = \mu_{z\tau}$. Conversely if $\alpha_1(y, t, T), \alpha_2(z, t)$ is a Nash equilibrium that satisfies the time invariant short-run player condition that the short-run player equilibrium beliefs $\alpha_1(z, t) = \alpha_1(z), \mu_{z\tau}(t) = \mu_{z\tau}$ and equilibrium strategy $\alpha_2(z, t) = \alpha_2(z)$ then $(\alpha_1(z), \alpha_2(z), \mu_{z\tau})$ is a time invariant equilibrium.

Hereafter by equilibrium we mean time invariant equilibrium.
Short-run Pure Equilibria

**Theorem:** For given $V, Q$ there exists an $\epsilon > 0$ such that for $\epsilon \min\{\pi, 1 - \pi\} > \epsilon > 0$ there is a unique short-run pure equilibrium. It is a strict Nash equilibrium, and in particular the long-run player also uses a pure strategy. The short-run player enters only on the good signal. There are three mutually exclusive types of equilibria depending on $c$ each corresponding to a different normal type long-run player pure strategy:

i. [bad] if $c > \delta$ the normal type never invests

ii. [trap] if $\delta > c > \frac{\pi}{1 - \delta + \delta \pi}$ the normal type invests only on the good signal

iii. [good] if $\frac{\pi}{1 - \delta + \delta \pi} > c$ the normal type always invests
Short-Run Player Behavior

assumption that $\epsilon$ is small means that types are highly persistent so the short-run player does not put much weight on the possibility of the type changing

signal $0$ indicates either a bad type or a normal type who will not invest if entry is not anticipated

• so short-run player should not enter in the face of bad signal

signal $1$ indicates either a good type or a normal type who will invest if entry is anticipated

• so short-run player should enter in the face of a good signal.
**No Signal**

can infer that the previous short-run player chose not to enter

- previous short-run player must have received the bad signal or was in the same boat with the signal $\mathcal{N}$

less decisive than the signal $0$ the signal $\mathcal{N}$ also indicates past bad behavior by the long-run player

- so staying out is a good idea
Long Run Player Optimality

invest when entry is anticipated? invest when entry is not anticipated? Changes chances of establishing a reputation...

modified problem for a long run player deciding whether to invest

a cost $c$ can be incurred resulting in probability $p$ of successfully establishing a good reputation and gaining $1 - c$ in the future

expected average present value of the gain from investment is

$$\Gamma = -(1 - \delta)c + \delta p(1 - c) + \delta (1 - p)\Gamma$$

or

$$\Gamma = \frac{\delta p(1 - c) - (1 - \delta)c}{1 - \delta (1 - p)}.$$

23
To Invest or not to Invest?

\[ \Gamma \text{ negative that is } \delta p(1 - c) < (1 - \delta)c \]

best not to invest and conversely

information is revealed immediately \( p = 1 \).

not invest for \( \delta < c \).

standard case, corresponding to part (i) of the Theorem in which the long-run player is impatient and does not find it worthwhile to give up \( c \) for a future gain of \( 1 - c \)

investment will only take place only occasionally during beneficial events when the good type finds it optimal to invest for non-reputational reasons.
To Invest

\[ \delta > c \] it is worth it to maintain a reputation when the short-run player enters as indeed in this case \( p = 1 \)

is also worth it to invest when the short-run player does not enter? Here \( p = \pi \)

the condition for that investment is as given in (ii) and (iii)

\[ c < \frac{\pi}{1 - \delta + \delta \pi} \]

corresponds to the “usual” reputational case: always invest

occasionally an adverse event occurs and the bad type finds it unprofitable to invest so investment does not take place until another normal or good type arrives
new case and the interesting case is case (ii)
\[ \delta > c \] interest rate is high enough to maintain a reputation, but
\[ c > \frac{\pi}{1 - \delta + \delta \pi} \] not worth it to try to acquire a reputation

strong history dependence.

two very different situations

a history of good signals investment, good reputation and have a wealthy and satisfactory life with an income of \( 1 - c \).

history in which the last signal was bad no investment a deservedly bad reputation, and have an impoverished life with an income of 0.
History Dependence

only different between these normal types is an event that took place in the far distant past: was the last event adverse or beneficial?

adverse and beneficial events, rare as they are, cast a very long shadow.

after a beneficial event there will be many lives of prosperous normal types – indeed until an adverse event occurs

following an adverse event normal types will be mired in the reputation trap until they are fortunate enough to have a beneficial event

for example, an outside threat that causes people to pull together (a beneficial event) may have very long-term consequences
Mixed Equilibria

Corollary: For given $V, Q$ there exists an $\epsilon > 0$ such that for $\epsilon \pi^2 (1 - \pi) > \epsilon > 0$ and

i. [bad] If $c > \delta$ then there is a unique equilibrium and there is no investment by the normal type.

ii. [trap] If $\delta > c > \delta \max \left\{ \frac{\pi}{1 - \delta + \delta \pi}, \frac{1}{1 + \delta (1 - \pi)} \right\}$ there is a unique equilibrium and it is a reputation trap.

iii. [good] If $\pi > (1 - \delta)/\delta$ and

$$\delta \frac{\pi}{1 - \delta + \delta \pi} > c > \delta \frac{1}{1 + \delta (1 - \pi)}$$

there is a unique equilibrium and there is always investment by the normal type.
Existence of Mixed Equilibria

Theorem: There exists fundamental bounds (not depending on $Q$, $\epsilon$) $\mu < 1$ and $\epsilon > 0$ such that for any $Q$ with $\mu_n \geq \mu$ if $\epsilon \mu \pi^2 (1 - \pi) > \epsilon > 0$ and

$$c < \delta \frac{1}{1 + \delta (1 - \pi)}$$

there is at least one single-mixing and one double-mixing equilibrium and no other type of mixed equilibrium. For each type of equilibrium there is a unique value of $\alpha_2(1)$. Moreover, for $z = 1$ in the single mixing case and $z \in \{N, 1\}$ in the double-mixing case the equilibrium value(s) of $\alpha_1(z)$ satisfies

$$|\alpha_1(z) - \bar{B}| \leq \frac{1 - \mu_n}{1 - \bar{\mu}}.$$
Mixed versus Pure

pure equilibria has three properties:

• the signal is informative for the short-run player
• reputation is valuable
• normal type of long-run player remains stuck in either a good or bad situation

mixed equilibria are quite different:

• the signal is uninformative for the short-run player
• reputation is not valuable,
• normal type of long-run player transitions back and forth between all the states
How to Get Out?

if Southern Italy is caught in a reputation trap, what might the central government of Italy or the EU do to help?

subsidize the cost of investment?

- $c$ is low enough then investment even with the bad signal will be profitable and - eventually - the trap will be escaped

- welfare analysis shows this is not a good idea (although of course it is what they do...)

long-run player already has the possibility of making the investment and finds it not worth while; if the money designated for an investment subsidy was instead given to the long-run player the long-run player would choose not to spend it on investment – and would be strictly better off
Buy Ferrari (race cars...) 

if $\pi$ could be increased it would be much easier to escape the reputation trap

an outside agency might have an advantage over the long-run agent having, perhaps, greater influence on outsiders and information flow to outsiders

sponsor large events such as a World Cup or the Olympics

by bringing large numbers of outsiders a cultural change is publicized – and that possibility increases the incentive for the change

one reason cities and regions compete for these events is precisely in hopes of obtaining favorable publicity
**Has it Ever Worked?**

to escape a reputation trap the investment must actually take place –

- Olympics in Athens in 2004 or in Rio in 2016 simply confirmed what everybody already believed about those cities

should also be emphasized that to be effective the increase in $\pi$ must be large enough - it must cross the threshold for which it becomes profitable to invest on the bad signal

possible positive examples

- Olympics in Barcelona in 1992 and the World Exposition in Chicago in 1893:

at the moment no satisfactory empirical analysis of these events exists