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
History Matters

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?

\[ a_1 \in \{0, 1\} \] with 1 meaning to invest and the cost being \( ca_1 \) 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, $N$

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 $N$

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

players are only aware of events that occur during their lifetime. short-run players being a short-lived mass are unable to coordinate on a mixed strategy.

- a strategy for a short-run player is a deterministic choice of action $\alpha_2(z) \in \{0, 1\}$ as a function of the beginning of period signal

long-run player observes finite histories $h$ of lifetime events and chooses a probability of investment $\alpha_1(h) \in [0, 1]$.

one common strategy used by all short-run players and one common strategy used by all long-run players – names and dates do not matter (different than Markov perfection because not all the information is payoff relevant)
Generic Cost

\[ c \notin \left\{ \delta, \frac{\delta}{2 - \pi}, \frac{\delta \pi}{1 - \delta + \delta \pi}, \frac{\delta \pi (\pi - \delta \pi)}{(1 - \delta \pi)(1 - \delta)} + \delta \pi (\pi - \delta \pi) \right\} \]
Beliefs and Equilibrium

given a strategy \( \alpha_2(z) \) of the short-run player the long-run player faces a Markov decision problem.

each period given the signal \( z \) a short-run player action \( \alpha_2(z) \) will result

optimization problem faced by the long-run player depends only on that action

\( V(a_2) \) denote the corresponding expected average value of utility
Bellman Equation

First period utility is $a_2 - ca_1$

Probability $\delta$ the game continues

Probability of the next signal $P(z'|z, a_1)$

- $P(1|z, 1) = P(0|z, 0) = \alpha_2(z) + (1 - \alpha_2(z))\pi$
- $P(N|z, a_1) = (1 - \alpha_2(z))(1 - \pi)$

Bellman equation

$$V(a_2) = \max_{a_1} (1 - \delta) [a_2 - ca_1] + \delta \sum_{z'} P(z'|z, a_1) V(\alpha_2(z'))$$

For small $\epsilon$ this problem has a unique solution
Purity of the Long-Run Player

as the Markov decision problem has a unique solution it is necessarily a pure Markov strategy

• we may assume that the long-run player chooses \( \alpha_1(z) \in \{0, 1\} \) depending only on the signal

simplifies the analysis of short-run player optimality.
The Ergodic Distribution

both the long-run and short-run players are following pure Markov strategies

hence joint distribution of types and signals $\mu_{zt}(t)$ follows a Markov process

short-run player chooses to go in $\alpha_2(z) = 1$ for all $z$ then the signal $N$ cannot occur; with this exception every state is reachable with positive probability from every other state so the Markov process is ergodic on the relevant state space

any initial beliefs therefore $\mu_{zt}(t)$ converge to a unique limit $\mu_{zt}$
**Short-Run Player Beliefs**

short-run player has no knowledge of time so the unique ergodic limit is taken to be the beliefs of the short-run player prior to observing the signal.
Equilibrium

consists of strategies \( \alpha_1(z), \alpha_2(z) \)

the unique ergodic beliefs \( \mu_{z_T} \) generated by those strategies

such that

- long-run player is playing a best response to the short-run player strategy
- short-run player is playing a best response to the ergodic beliefs and the signal


**Main Theorem**

For given $V, Q$ there exists an $\varepsilon > 0$ such that for $\varepsilon \min\{\pi, 1 - \pi\} > \varepsilon > 0$ there is a unique equilibrium, it is strict, and the short-run player enters only on the good signal. There are three mutually exclusive types of equilibria

i. if $\delta < c$ the long-run player never invests, otherwise

ii. if $\pi < \frac{(c - \delta c)}{(\delta - \delta c)}$ the long-run player invests only on the good signal

iii. if $\pi > \frac{(c - \delta c)}{(\delta - \delta c)}$ the long-run player always invests.

Note that the boundary cases are ruled out by the generic cost assumption.
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 $N$

less decisive than the signal 0 the signal $N$ 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

$$\Gamma = \frac{\delta p(1 - c) - (1 - \delta)c}{1 - \delta(1 - p)}.$$
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 investment is that given in (ii) and (iii)

\[ \pi \leftrightarrow \frac{(c - \delta c)}{(\delta - \delta c)} \]

if \( \pi \) is high enough that positive news spreads quickly then it is worth investing even when the short-run player does not enter.

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
The Reputation Trap

new case and the interesting case is case (ii)

\[ \delta > c \] interest rate is high enough to maintain a reputation, but
\[ \pi < \frac{(c - \delta c)}{(\delta - \delta c)} \] 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
The Role of Behavioral Types

no behavioral types:

static Nash equilibrium – always stay out and never invest – is an equilibrium

• case (i) only equilibrium

• higher discount factors both the case (ii) and case (iii) strategies are Nash equilibria

only one that is subgame perfect is the case (ii) equilibrium in case (ii)
presence of good types eliminates the static Nash equilibrium once the discount factor is high enough

bad types, however, are key in selecting between the (ii) and (iii) type equilibria
**Bad Types**

presence of behavioral types insures that the ergodic distribution is unique and that all signals (except possibly $N$) are present

acts somewhat like trembles.

non-subgame perfect type (iii) equilibrium is eliminated in case (ii)

type (ii) equilibrium eliminated in case (iii)

  • play must be optimal following a signal of no investment

type (iii) equilibrium in case (iii).

  • despite the fact that the normal types always invest it is optimal for the short-run player to stay out on a signal of no investment: this is because such a signal indicates a bad type
**Mixed Strategies**

few normal types of long-run players there is no mixed equilibrium

many normal types: can have a signal jamming equilibrium

• normal long-run type mixes and there are few behavioral types so
  the signal from last period – whatever it may be – is far more likely
  to come from a normal type who is mixing than a behavior type

• reputational arguments based on inferences about behavioral types
  break down

with longer memory, it is easier to distinguish a behavioral type from a
normal type who mixes
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