Catastrophe by Design in Population Games

Destabilizing Wasteful Locked-in Technologies

Stefanos Leonardos, Iosif Sakos, Costas Courcoubetis and Georgios Piliouras

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

10 YEARS IGNITING INNOVATION BY DESIGN

Phase Transitions in Societal Systems

Can we flip social behaviour?

Motivation

The adoption of innovative, socially-optimal technologies is usually hindered or stopped by wasteful lock-ins.

Various technological dilemmas over the years:

- TCP/IP vs. ATM
- IPv4 vs. IPv6
- Conventional vs. Electric Vehicles
- Proof-of-Work (PoW) vs. Proof-of-Stake (PoS)
- etc.

Bifurcation and Catastrophe Theory finds application in these cases.

Model I: Agents and Strategies

A population of agents

- Mass $K > 0$: total population size.
- Strategies: two available technologies, W (wasteful), and S (socially-optimal).
- Inherent cost: $\gamma > 0$ for W and 0 for S.
- Population states: $X = \{(x, 1-x) : x \in [0,1]\}$ where $x =$ fraction of adopters of the wasteful technology.

Model II: Value and Payoffs

Each technology creates value split among adopters

- Value V, Growth $\alpha > 1$:
  - $V_W = V(xK)^\alpha$ and $V_S = V((1-x)K)^\alpha$
- Payoff functions: equal share amongst all participating agents:
  - $u(W, x) = V_W \cdot (xK)^{-1} - \gamma = VK^{\alpha-1}x^{\alpha-1} - \gamma$
  - $u(S, x) = V_S \cdot ((1-x)K)^{-1} = VK^{\alpha-1}(1-x)^{\alpha-1}$
- For the purposes of this talk we restrict ourselves to the case $\alpha = 2$:
  - $u(W, x) = VKx - \gamma$
  - $u(S, x) = VK(1-x)$

An Evolutionary Game

Evolutionary game interpretation

$$P = \begin{array}{c} \\ W \\ S \end{array} \begin{array}{c} \begin{array}{cc} W & S \end{array} \\ \begin{pmatrix} VK - \gamma & -\gamma \\ 0 & VK \end{pmatrix} \end{array} \tag{G1}$$

Theorem

(G1) has three equilibria: (S) and (W) the monomorphic/pure states where everyone chooses S and W, respectively, and one mixed. The two pure equilibria are evolutionary stable, whereas the mixed one is unstable.

Population Dynamics

Q-Learning dynamics:

$$\dot{x} = x[\underbrace{u(W,x) - \bar{u}(x)}_{\text{Replicator Dynamics}} - T \cdot \underbrace{(x \ln x + (1-x)\ln(1-x))}_{\text{Entropy}}]$$

Where $\bar{u}(x) = xu(W,x) + (1-x)u(S,x)$

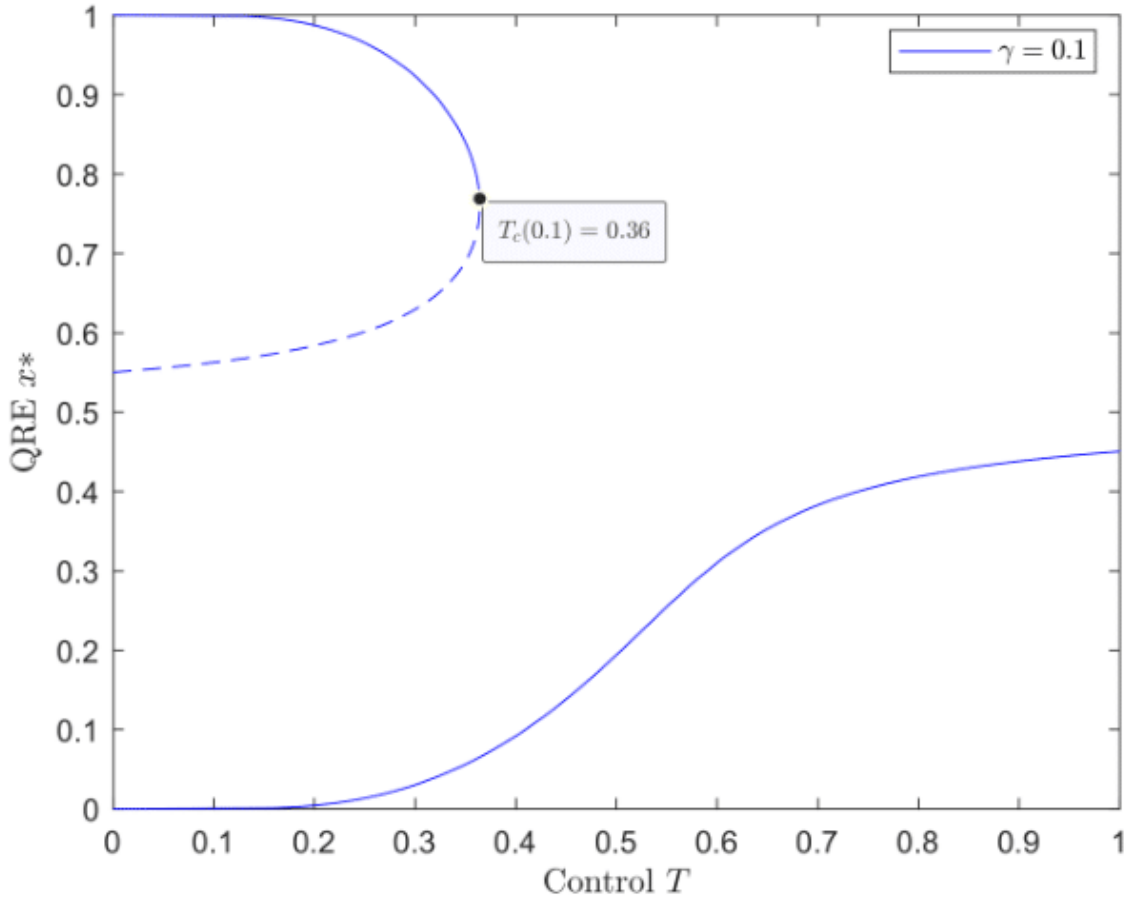
Quantal Response Equilibrium (QRE): The steady states of the system, i.e., $\dot{x} = 0$.

We can affect the agents' rationality by scaling the agents' utilities:

$$x[\frac{u(W,x)}{c} - \frac{\bar{u}(x)}{c} - T \cdot (x \ln x + (1-x)\ln(1-x))] = 0$$
$$\iff x[u(W,x) - \bar{u}(x) - cT \cdot (x \ln x + (1-x)\ln(1-x))] = 0$$

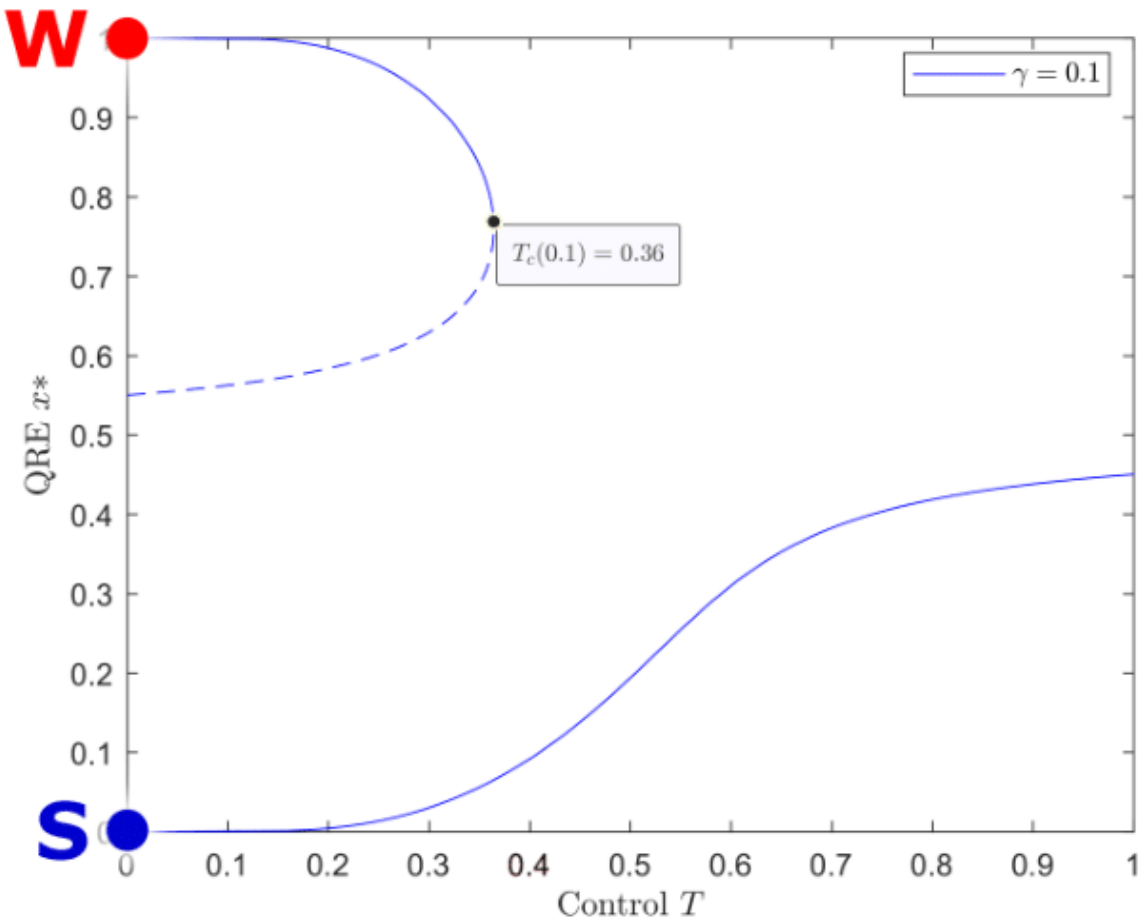
QRE Correspondence: Visually

In our case ($\alpha = 2$): $\dot{x} = x(1-x)[2x - (1+\gamma) - T\ln(\frac{x}{1-x})]$



Can we Flip the Population's Behaviour?

Can we flip the system from the bad equilibrium to the good one?



QRE Correspondence: Formally

Theorem

For any $\alpha > 1$ there exists a finite sequence of temperatures $T =< T_0, T_1, \cdots >$ such as starting from an initial state x_0 and performing the following procedure for each $T_i \in T$:

- *Scale the system's temperature at T_i, and*
- *Wait until the system converges to a QRE*

the system is going to converge to the desirable state $x = 0$ which corresponds to socially-optimal technology S.

We can reliably destabilize a wasteful equilibrium and converge to a socially-optimal equilibrium by introducing and removing taxes in the system.

Short Term Policy $\Longrightarrow$ **Long Lasting Effects**

QRE Correspondence: Proof I

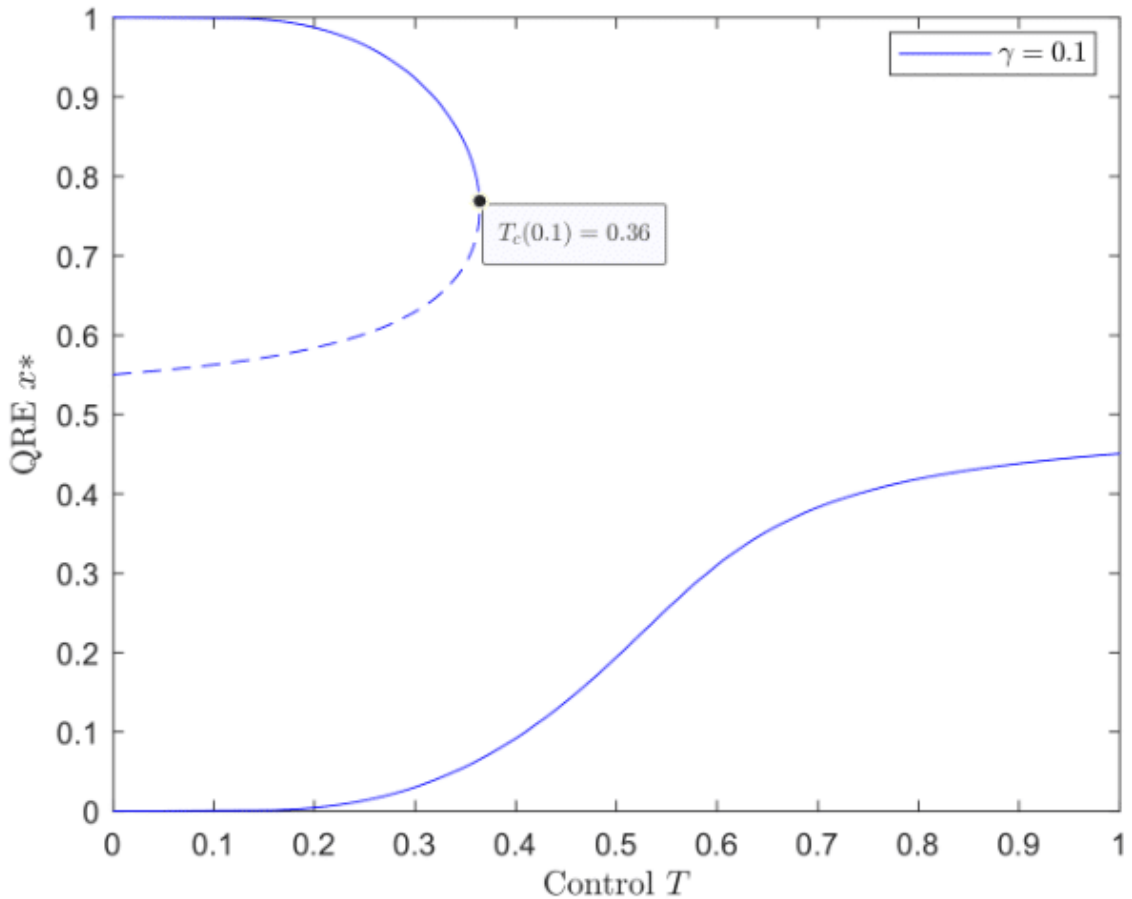
Sketch of the proof:

1. Show that for every fixed temperature $T \geq 0$, the system always converges to equilibrium points.
2. Design a sequence of temperatures such that as we move through the sequence of self-stabilizing systems the final system converges to the socially-optimal equilibrium.

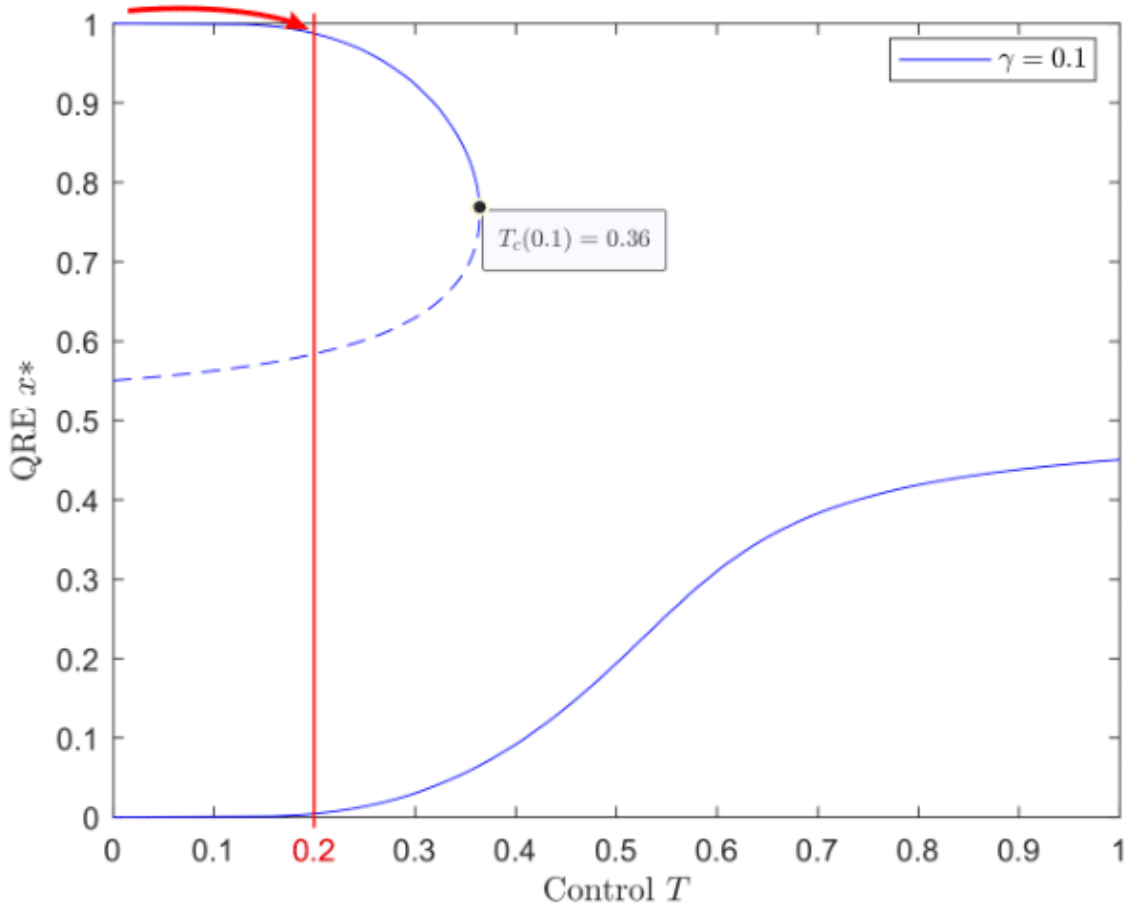
Step 1 is proved via Lyapunov Theory.

Step 2 is proved via Catastrophe or Bifurcation Theory.

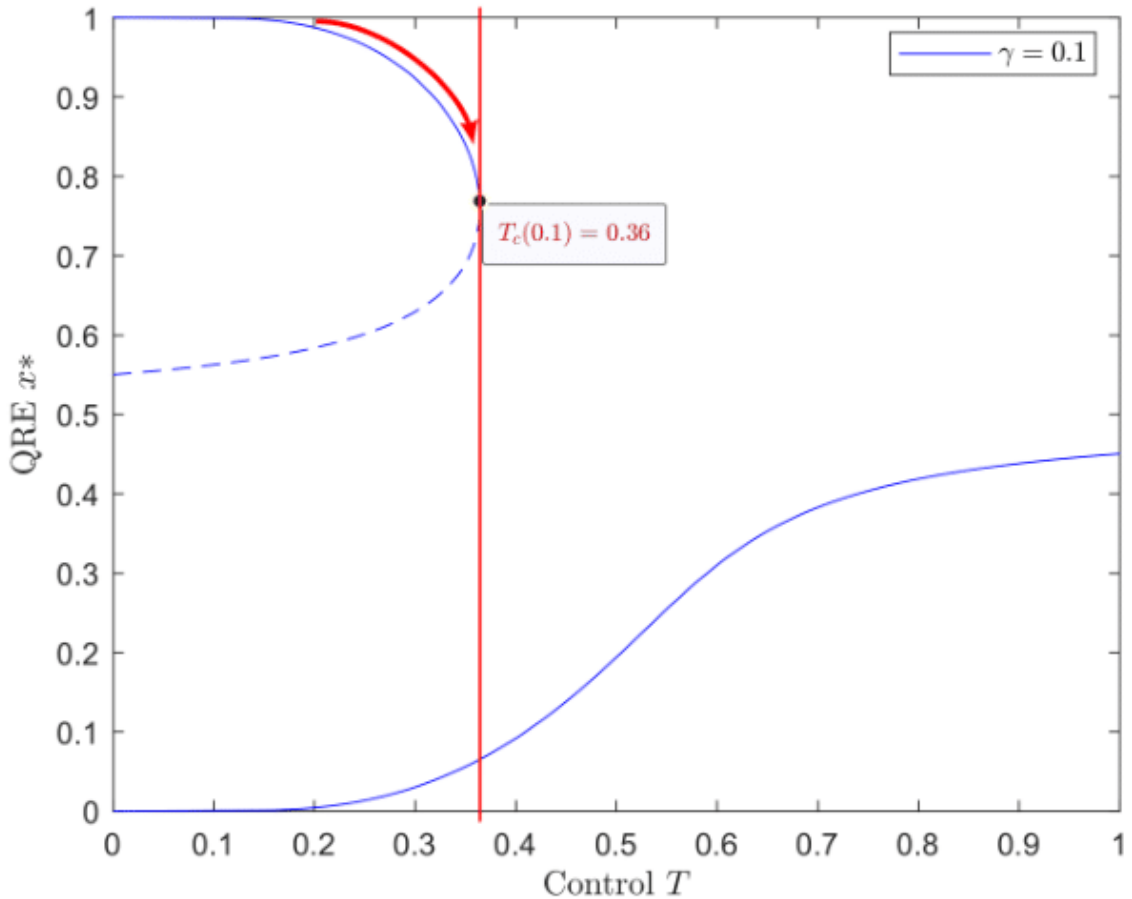
QRE Correspondence: Proof II



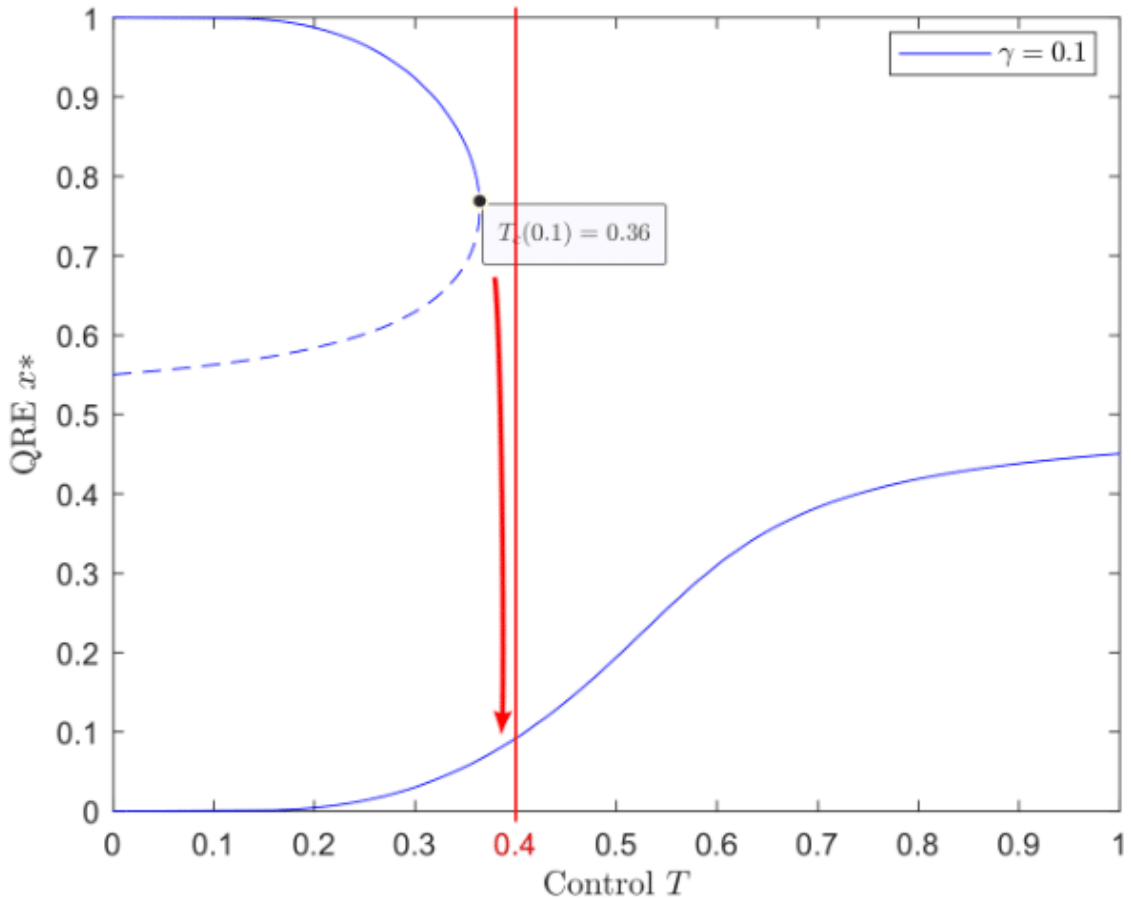
QRE Correspondence: Phase 1



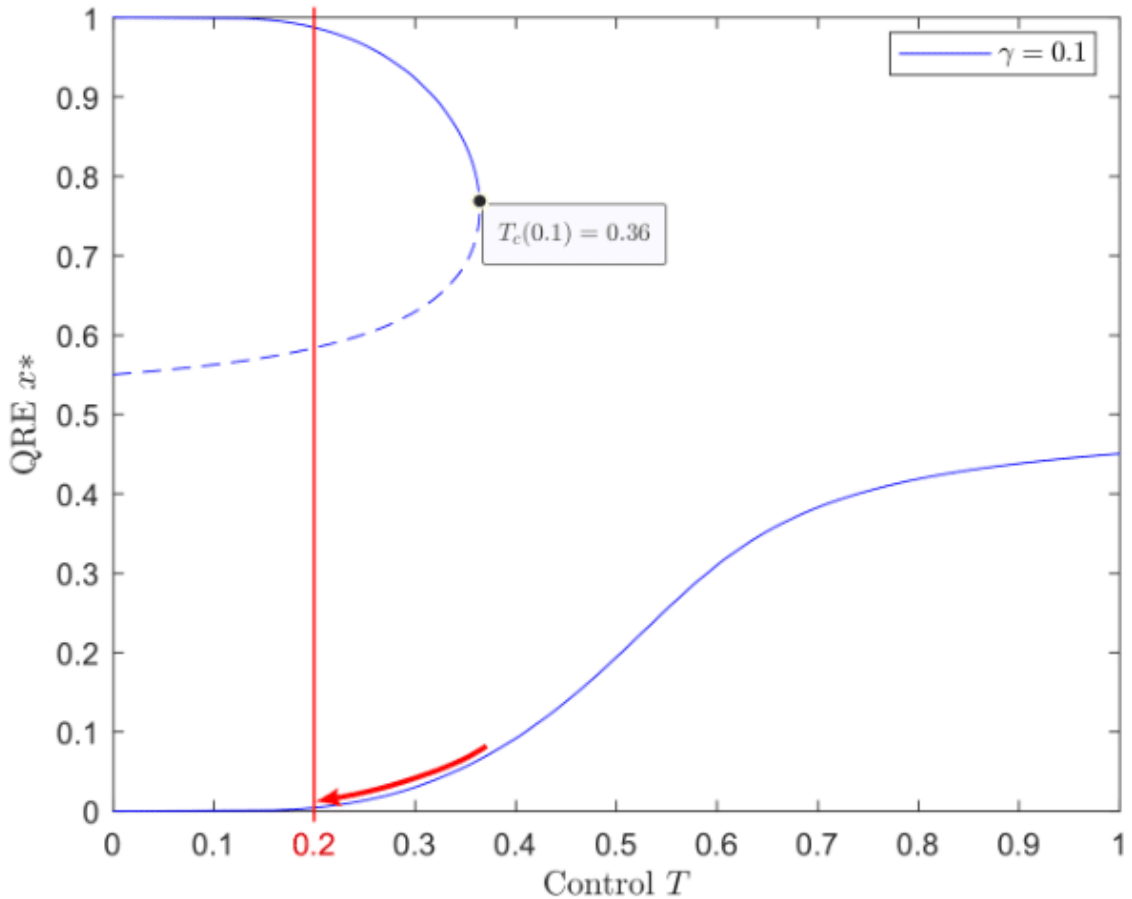
QRE Correspondence: Phase 2



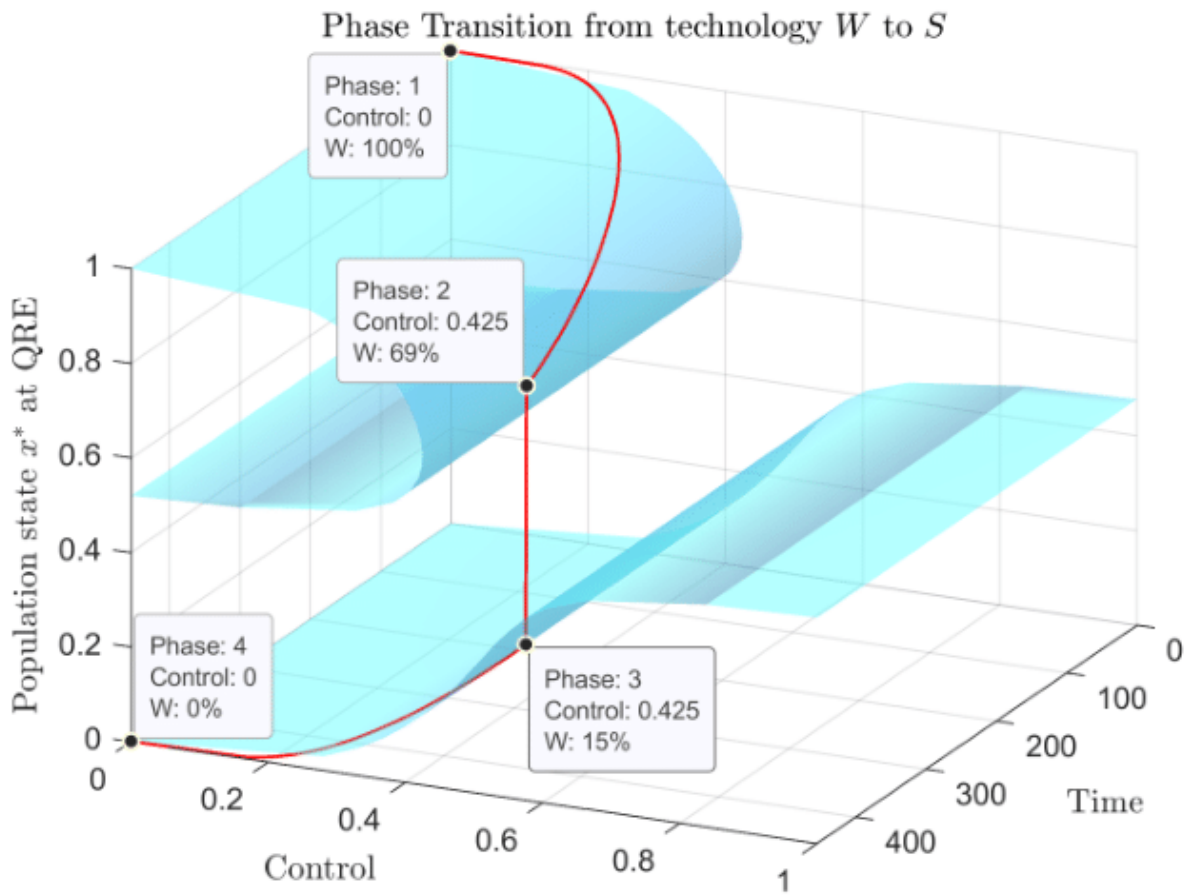
QRE Correspondence: Phase 3



QRE Correspondence: Phase 4



Designing a Catastrophe



Conclusion

Can we flip social behaviour?

Yes, we can!

HOW? CATASTROPHE BY DESIGN

- *Destabilize* ***bad equilibrium***
- *Induce* ***bifurcation*** *(# Equilibria change)*
- ***Phase transition*** *(jump to another branch of equilibria)*
- *Remove control to stabilize* ***good equilibrium***