Social preferences, positive expectations, and trust based cooperation

Ryan O. Murphy and Kurt Ackermann

Chair of Decision Theory and Behavioral Game Theory
ETH Zürich

ICSD June 24th, 2015

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rmurphy@ethz.ch

Overview

- Started with a discussion and a question:
  - How is SVO related to trust?

Overview

- Started with a discussion and a question:
  - How is SVO related to trust?
  - SVO is a necessary, but not sufficient, precursor to trust based cooperation.

Overview

I. PD game in a generalized sense
II. SVO- Social preferences
III. Beliefs- Positive expectations
IV. Trust based cooperation

Structure of the of the talk

PD game

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$T > R > P > S$$

$$2R > (T + S)$$

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$T > R > P > S$$

$$2R > (T + S)$$

Defining the interaction

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$T > R > P > S$$

$$2R > (T + S)$$

But many people choose to cooperate-- Even if the game is one shot, fully anonymous, and incentive compatible interaction.

Why?

Why would someone choose to cooperate?

Trust

Player 2

Player 1		C	D
	C	R, R	S, T
	D	T, S	P, P

$$T > R > P > S$$

$$2R > (T + S)$$

Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another.

Rousseau et al., 1998

Why would someone choose to cooperate?

Trust

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$T > R > P > S$$

$$2R > (T + S)$$

Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another.

Rousseau et al., 1998

...with the intention of improving collective outcomes.

Player 1 \ Player 2	C	D
C	R, R	S, T
D	T, S	P, P

$$T > R > P > S$$

$$2R > (T + S)$$

$$T = 1 \qquad S = 0$$

Player 2

Player 1		C	D
	C	R, R	S, T
	D	T, S	P, P

$$T > R > P > S$$

$$2R > (T + S)$$

$$T = 1 \qquad S = 0$$

$$R, P \in (0, 1)$$

Generalizing the PD game

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$T > R > P > S$$

$$2R > (T + S)$$

$$T = 1 \qquad S = 0$$

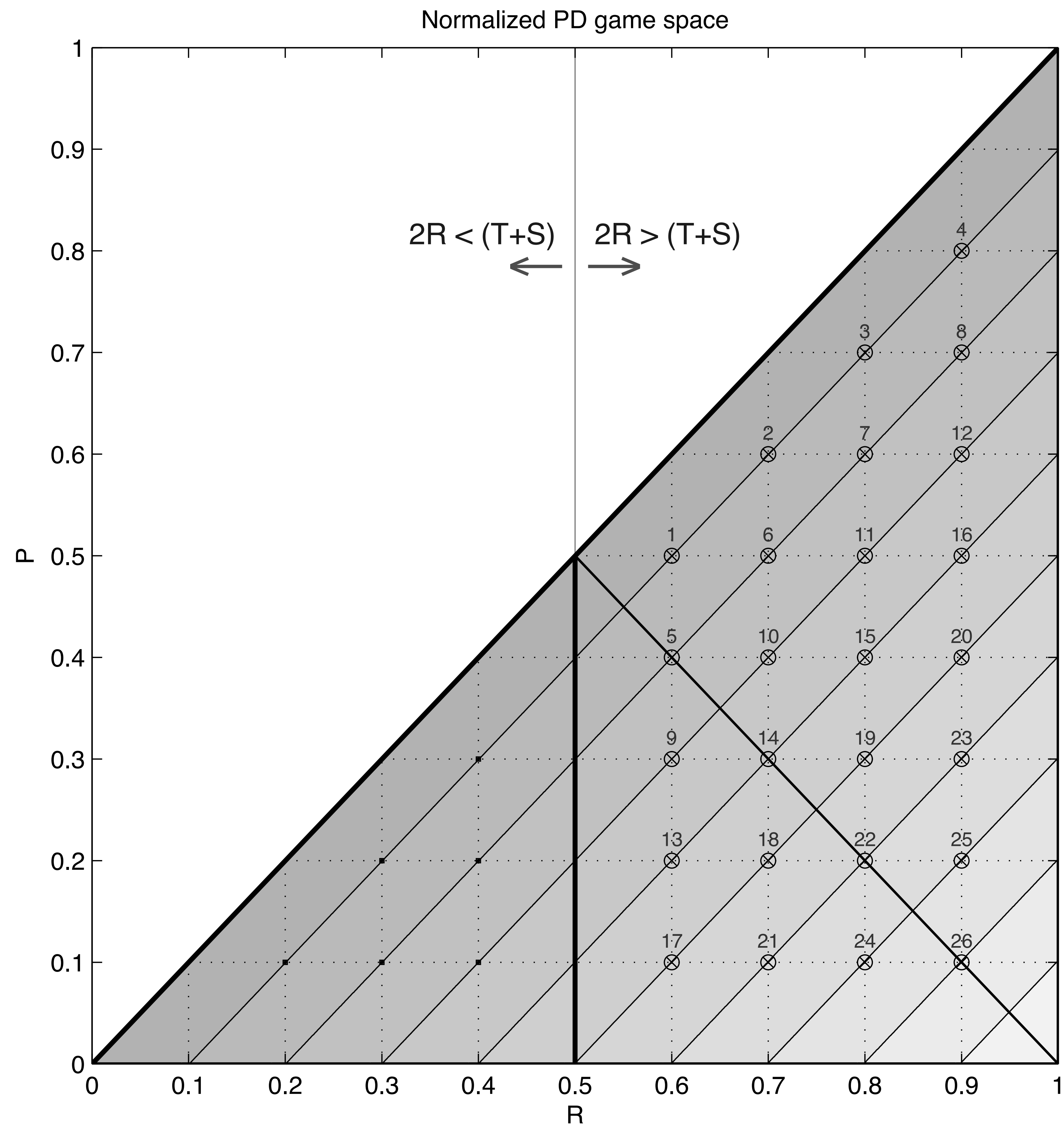
$$R, P \in (0, 1)$$

		Player 2	
		C	D
Player 1	C	.8, .8	0, 1
	D	1, 0	.6, .6

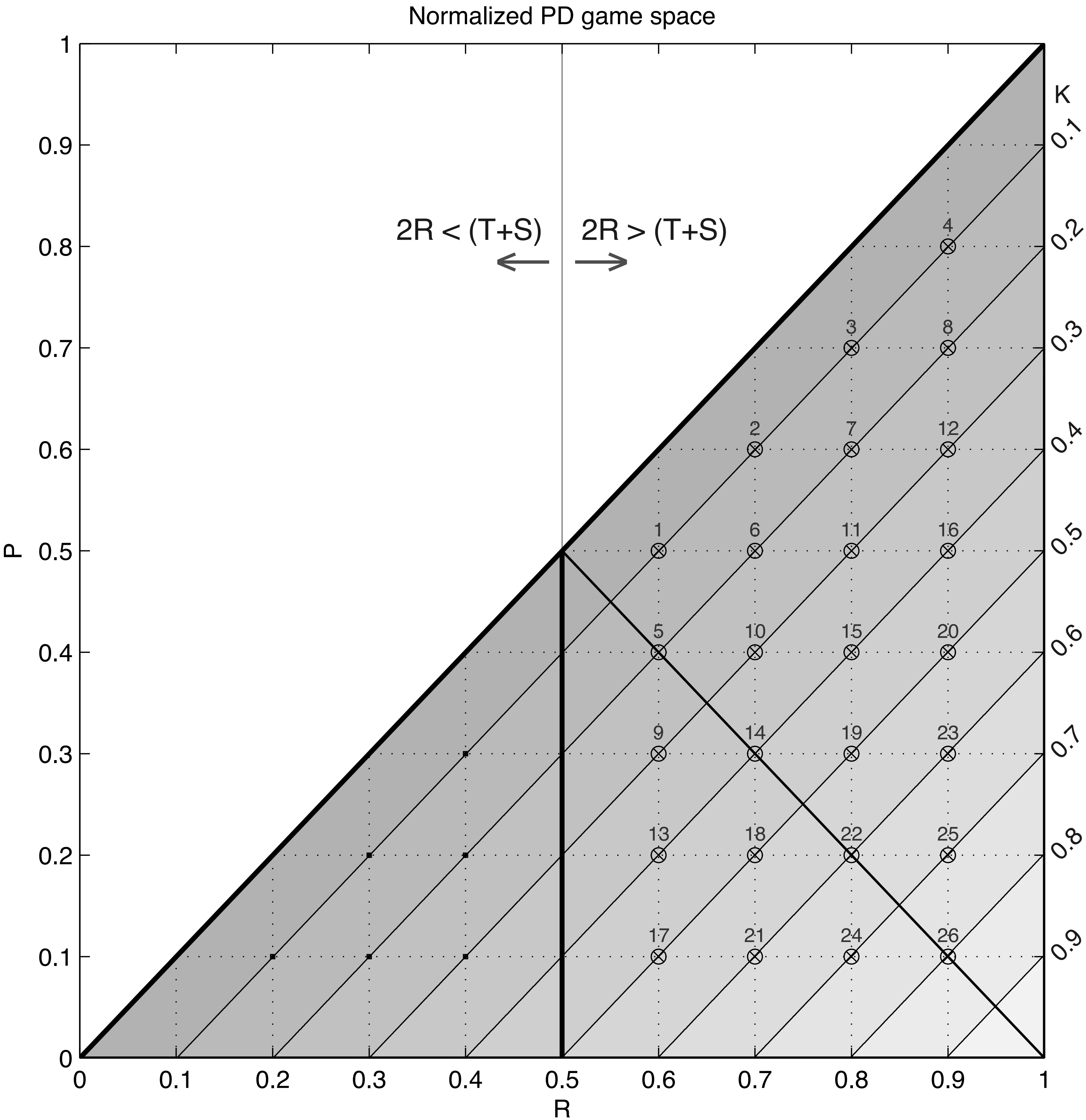
		Player 2	
		C	D
Player 1	C	.6, .6	0, 1
	D	1, 0	.4, .4

		Player 2	
		C	D
Player 1	C	.9, .9	0, 1
	D	1, 0	.1, .1

PD game	T	R	P	S
1	1	0.6	0.5	0
2	1	0.7	0.6	0
3	1	0.8	0.7	0
4	1	0.9	0.8	0
5	1	0.6	0.4	0
6	1	0.7	0.5	0
7	1	0.8	0.6	0
8	1	0.9	0.7	0
9	1	0.6	0.3	0
10	1	0.7	0.4	0
11	1	0.8	0.5	0
12	1	0.9	0.6	0
13	1	0.6	0.2	0
14	1	0.7	0.3	0
15	1	0.8	0.4	0
16	1	0.9	0.5	0
17	1	0.6	0.1	0
18	1	0.7	0.2	0
19	1	0.8	0.3	0
20	1	0.9	0.4	0
21	1	0.7	0.1	0
22	1	0.8	0.2	0
23	1	0.9	0.3	0
24	1	0.8	0.1	0
25	1	0.9	0.2	0
26	1	0.9	0.1	0



PD game	T	R	P	S	K
1	1	0.6	0.5	0	0.10
2	1	0.7	0.6	0	0.10
3	1	0.8	0.7	0	0.10
4	1	0.9	0.8	0	0.10
5	1	0.6	0.4	0	0.20
6	1	0.7	0.5	0	0.20
7	1	0.8	0.6	0	0.20
8	1	0.9	0.7	0	0.20
9	1	0.6	0.3	0	0.30
10	1	0.7	0.4	0	0.30
11	1	0.8	0.5	0	0.30
12	1	0.9	0.6	0	0.30
13	1	0.6	0.2	0	0.40
14	1	0.7	0.3	0	0.40
15	1	0.8	0.4	0	0.40
16	1	0.9	0.5	0	0.40
17	1	0.6	0.1	0	0.50
18	1	0.7	0.2	0	0.50
19	1	0.8	0.3	0	0.50
20	1	0.9	0.4	0	0.50
21	1	0.7	0.1	0	0.60
22	1	0.8	0.2	0	0.60
23	1	0.9	0.3	0	0.60
24	1	0.8	0.1	0	0.70
25	1	0.9	0.2	0	0.70
26	1	0.9	0.1	0	0.80

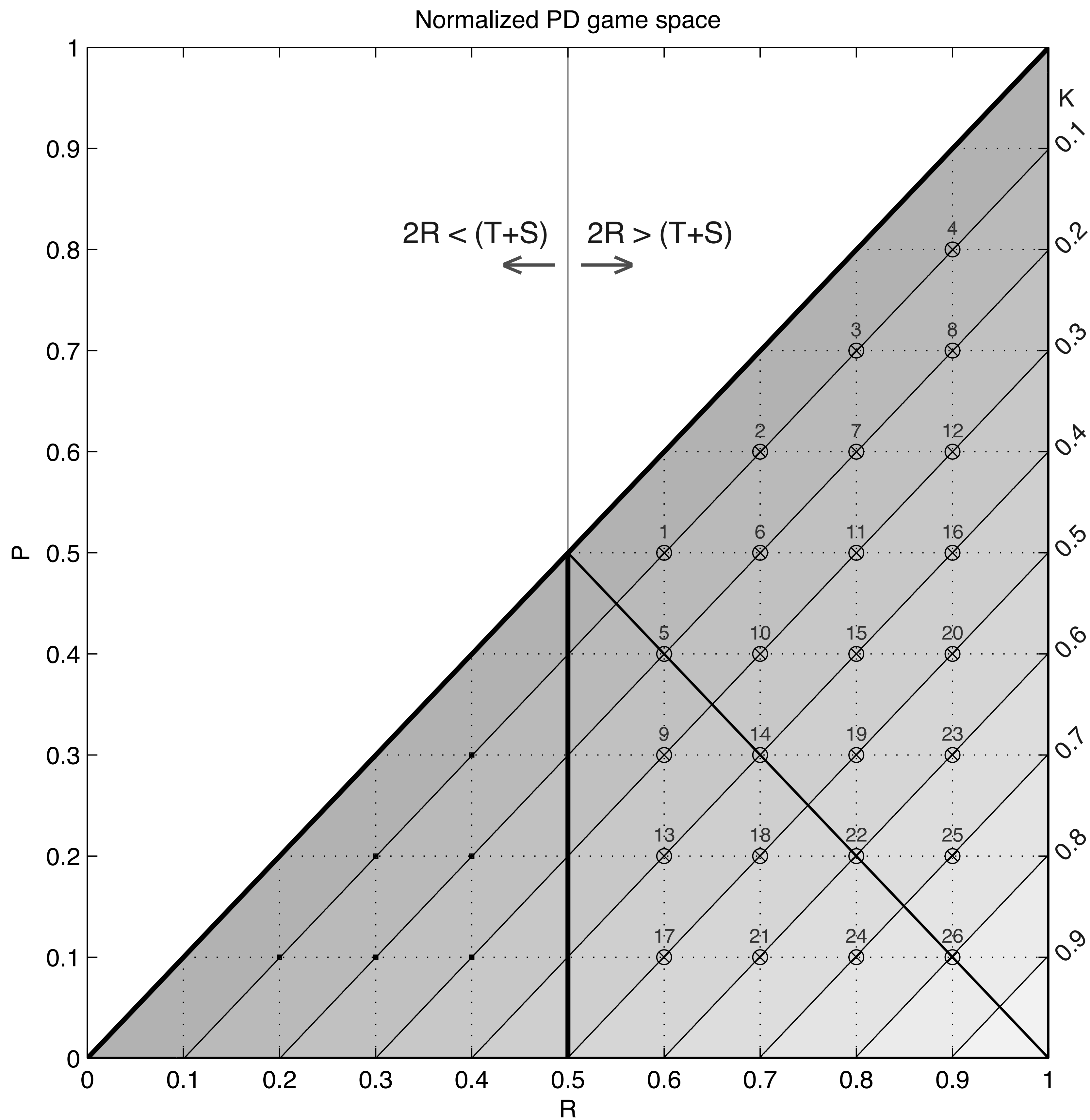


$$K = \frac{(R - P)}{(T - S)}$$

		Player 2 C	Player 2 D
Player 1	C	.9, .9	0, 1
	D	1, 0	.8, .8

		Player 2 C	Player 2 D
Player 1	C	.6, .6	0, 1
	D	1, 0	.4, .4

		Player 2 C	Player 2 D
Player 1	C	.9, .9	0, 1
	D	1, 0	.1, .1



$$K = \frac{(R-P)}{(T-S)}$$

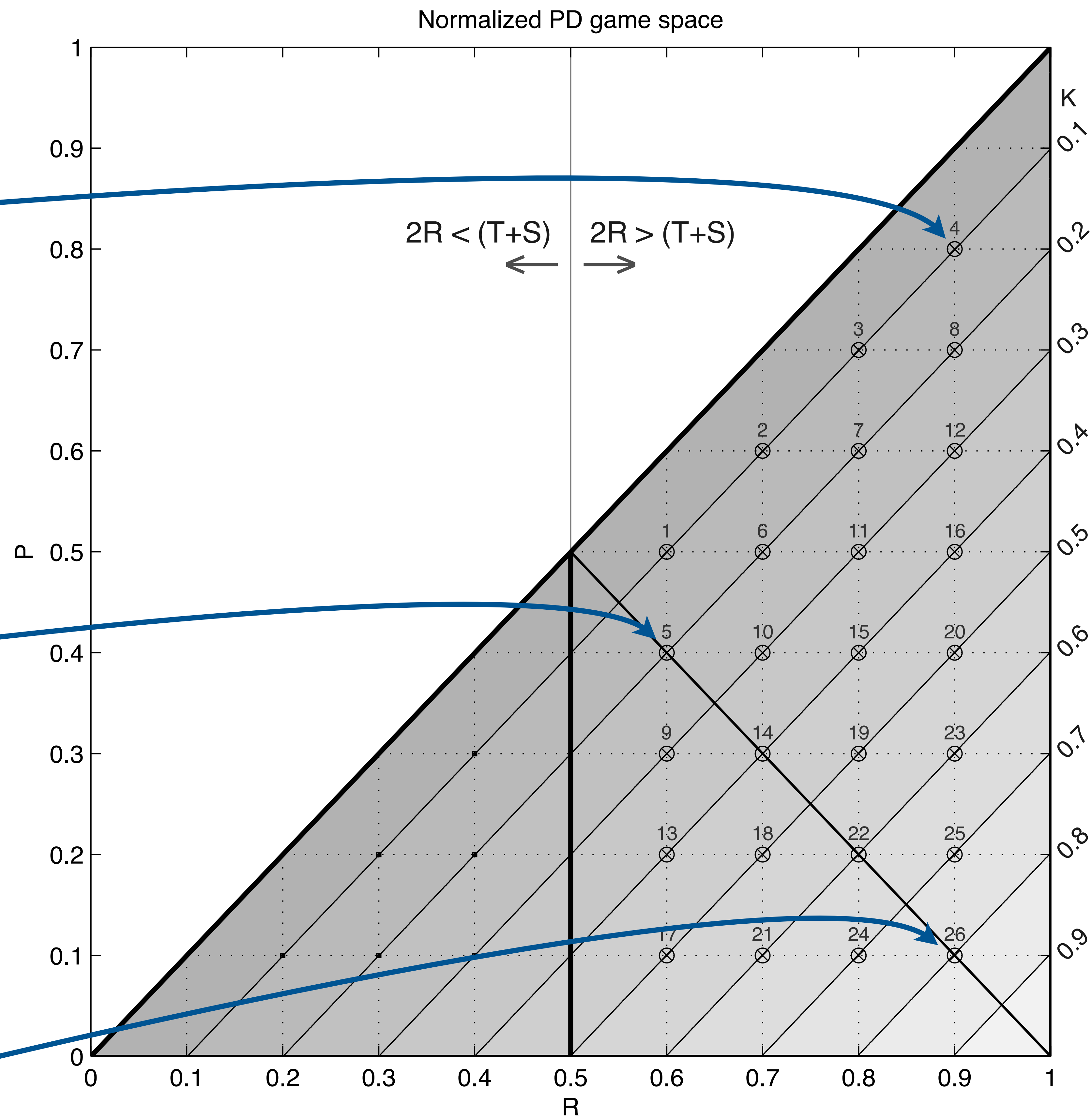
		Player 2 C	Player 2 D
Player 1	C	.9, .9	0, 1
	D	1, 0	.8, .8

More severe

		Player 2 C	Player 2 D
Player 1	C	.6, .6	0, 1
	D	1, 0	.4, .4

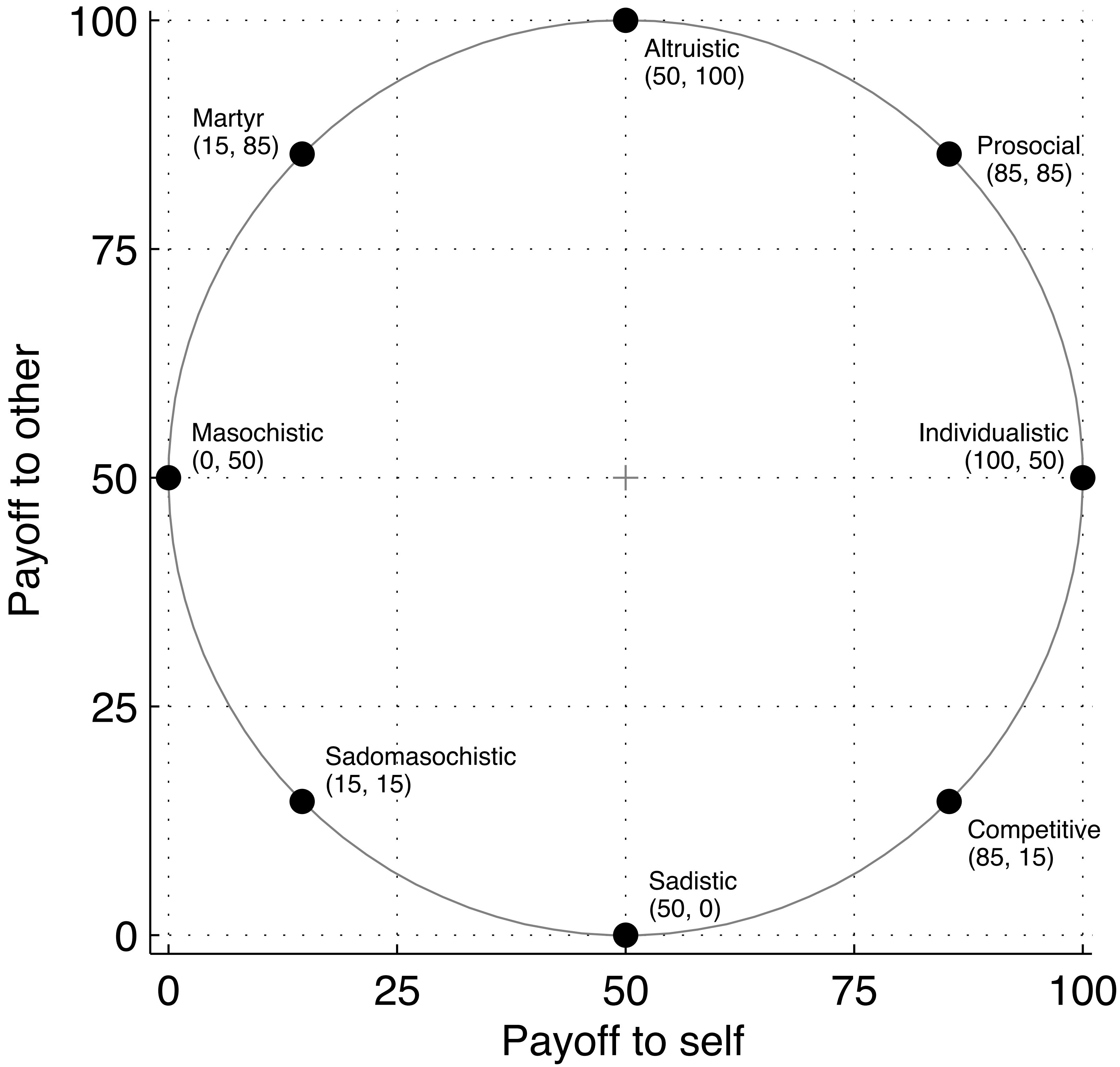
		Player 2 C	Player 2 D
Player 1	C	.9, .9	0, 1
	D	1, 0	.1, .1

Less severe



$$K = \frac{(R-P)}{(T-S)}$$

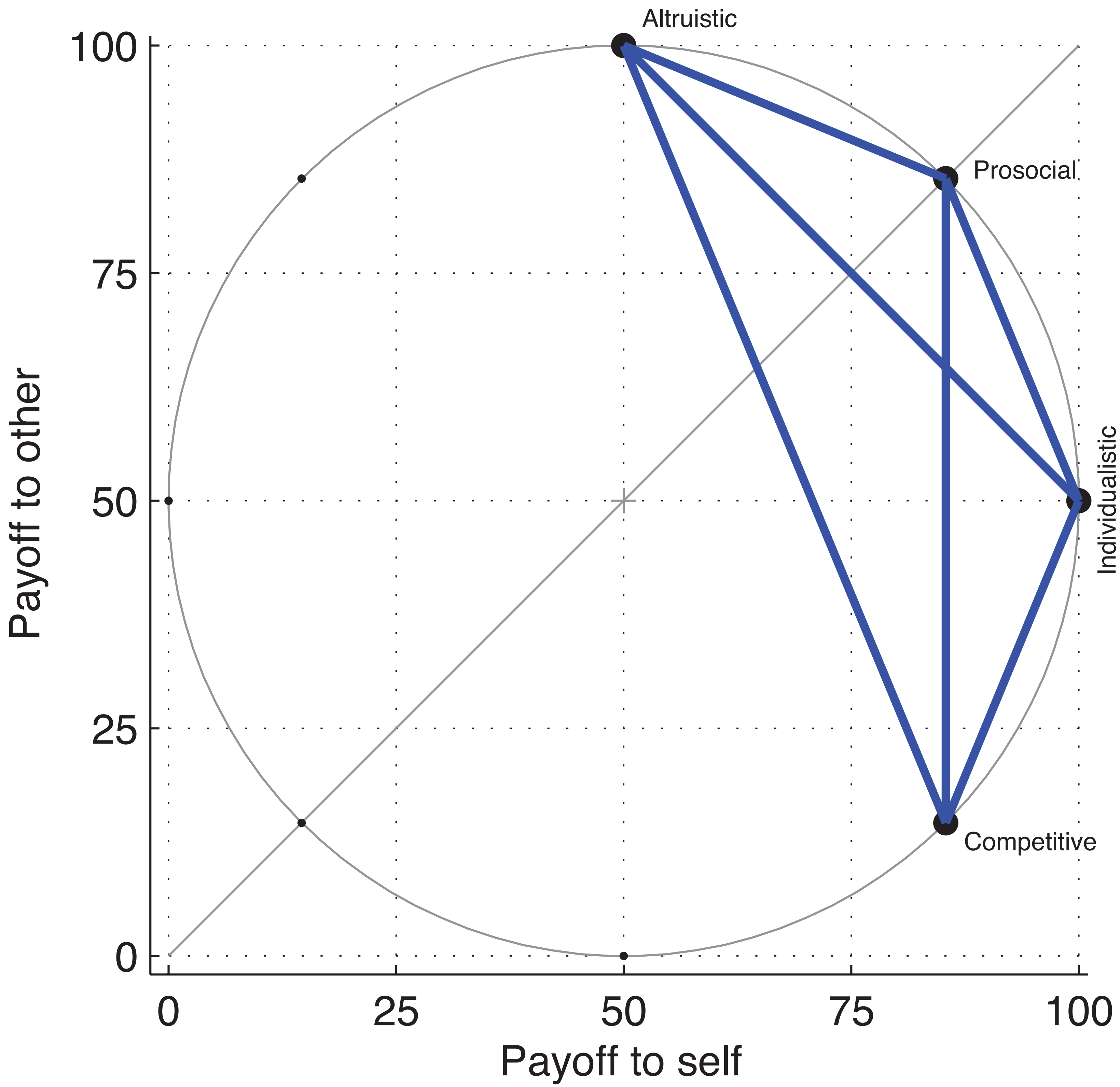
SVO



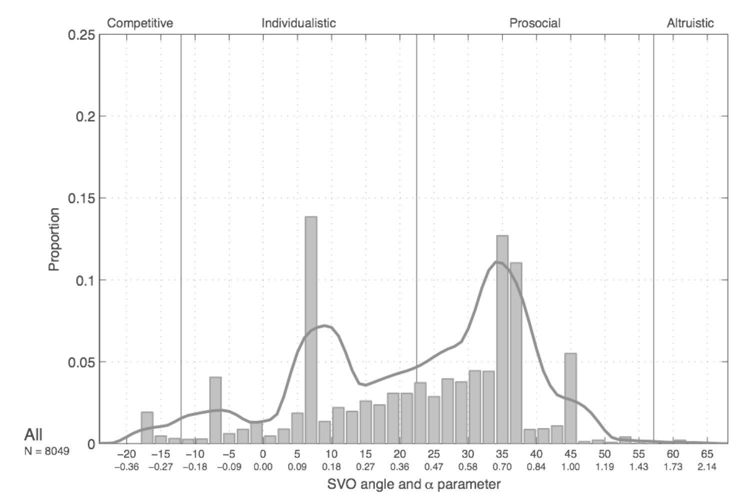
$$u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o$$

Edgeworth (1881), Pruitt (1967), Messick & McClintock (1968), Librand (1988), Van Lange (1999)

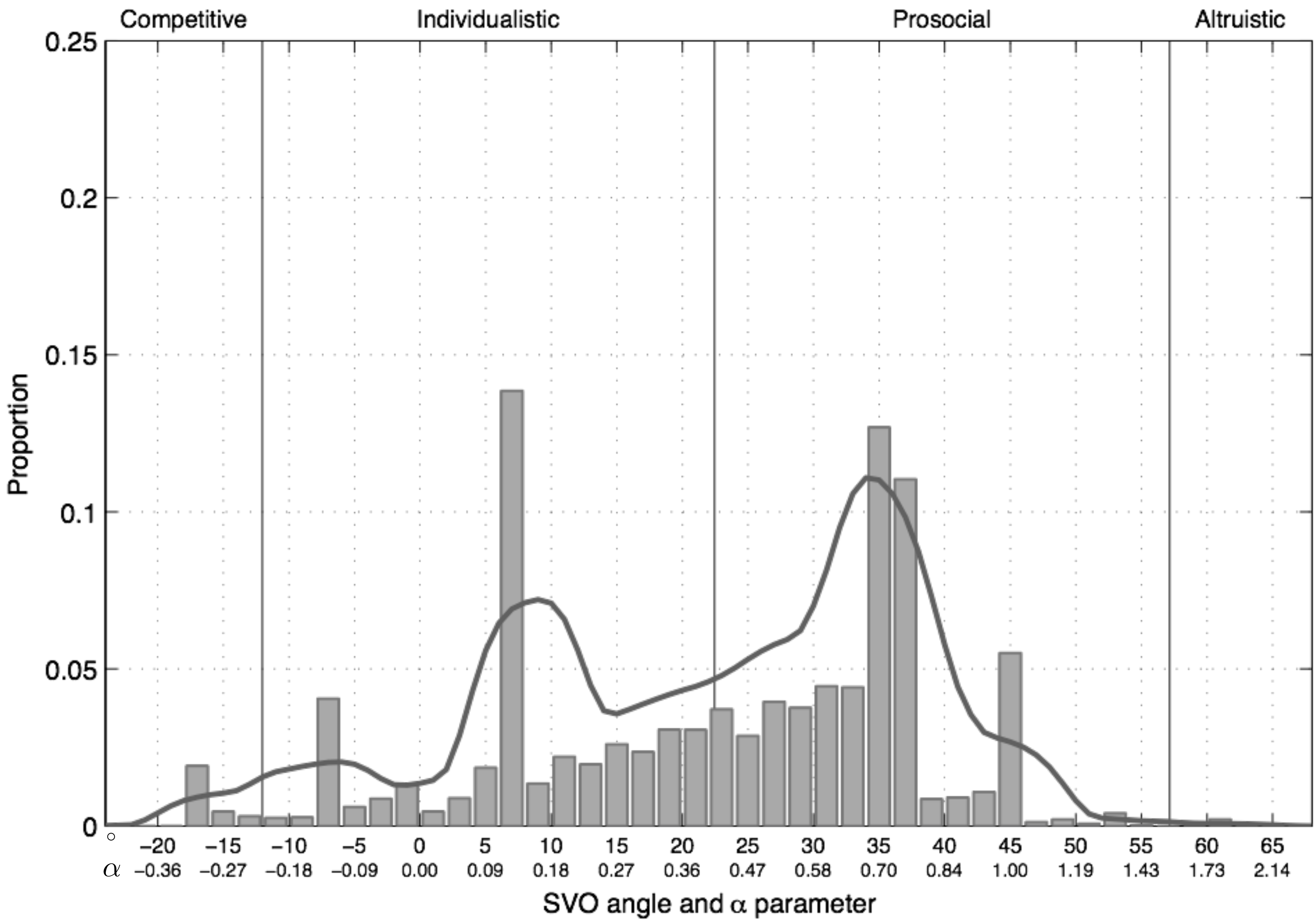
SVO



$$u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o$$



Slider measure



Social value orientation results

Beliefs

- How likely does a DM believe it is that the other player will choose to cooperate.
  - Zero if the DM is certain the other will defect
  - One if the DM is certain the other will cooperate
  - In-between values correspond to different degrees of belief about what the other player will choose

$$\beta \in [0, 1]$$

Integration

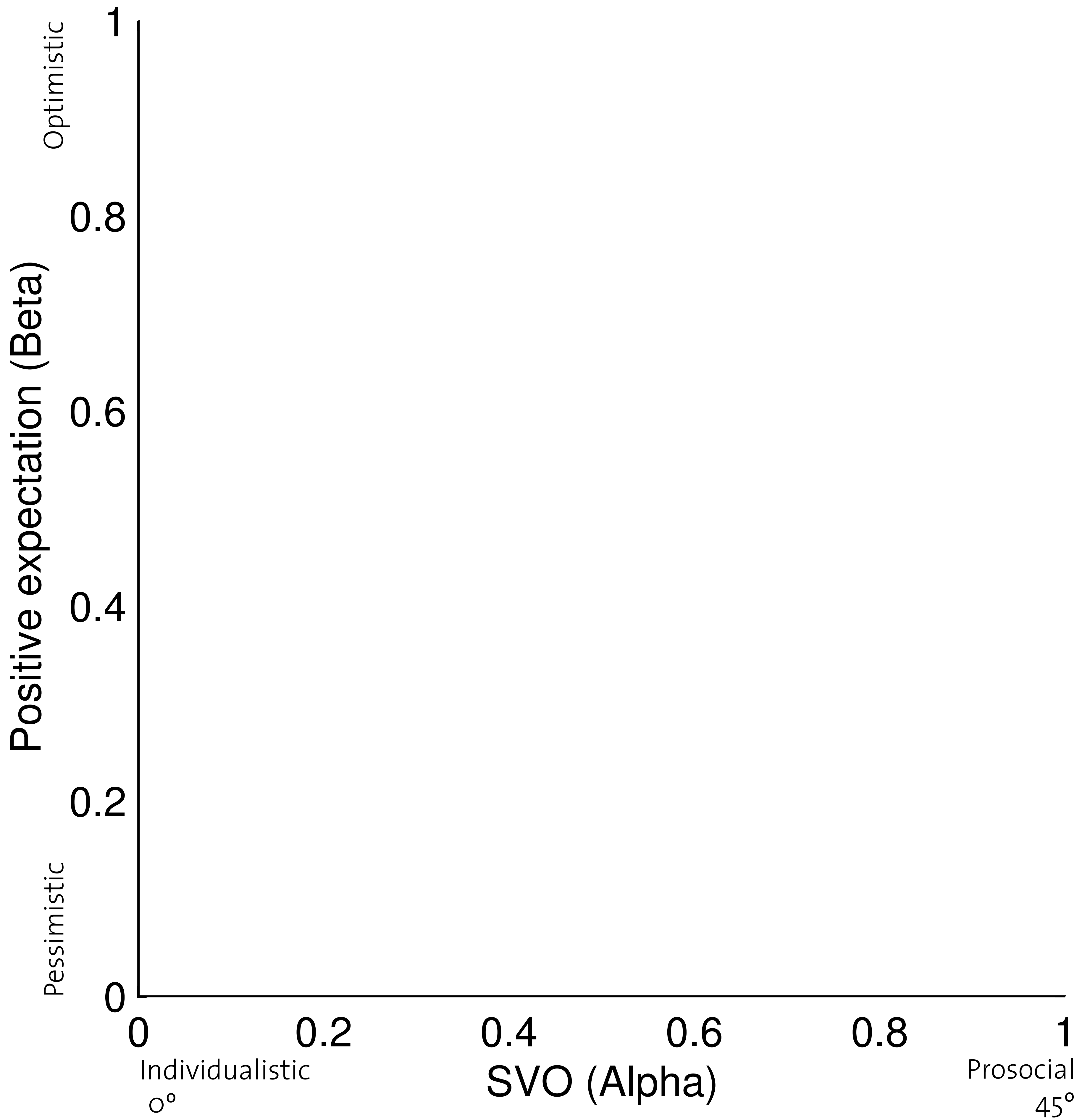
$\int$

- PD game in a generalized sense
- SVO- Social preferences
- Beliefs- Positive expectations
- Trust based cooperation

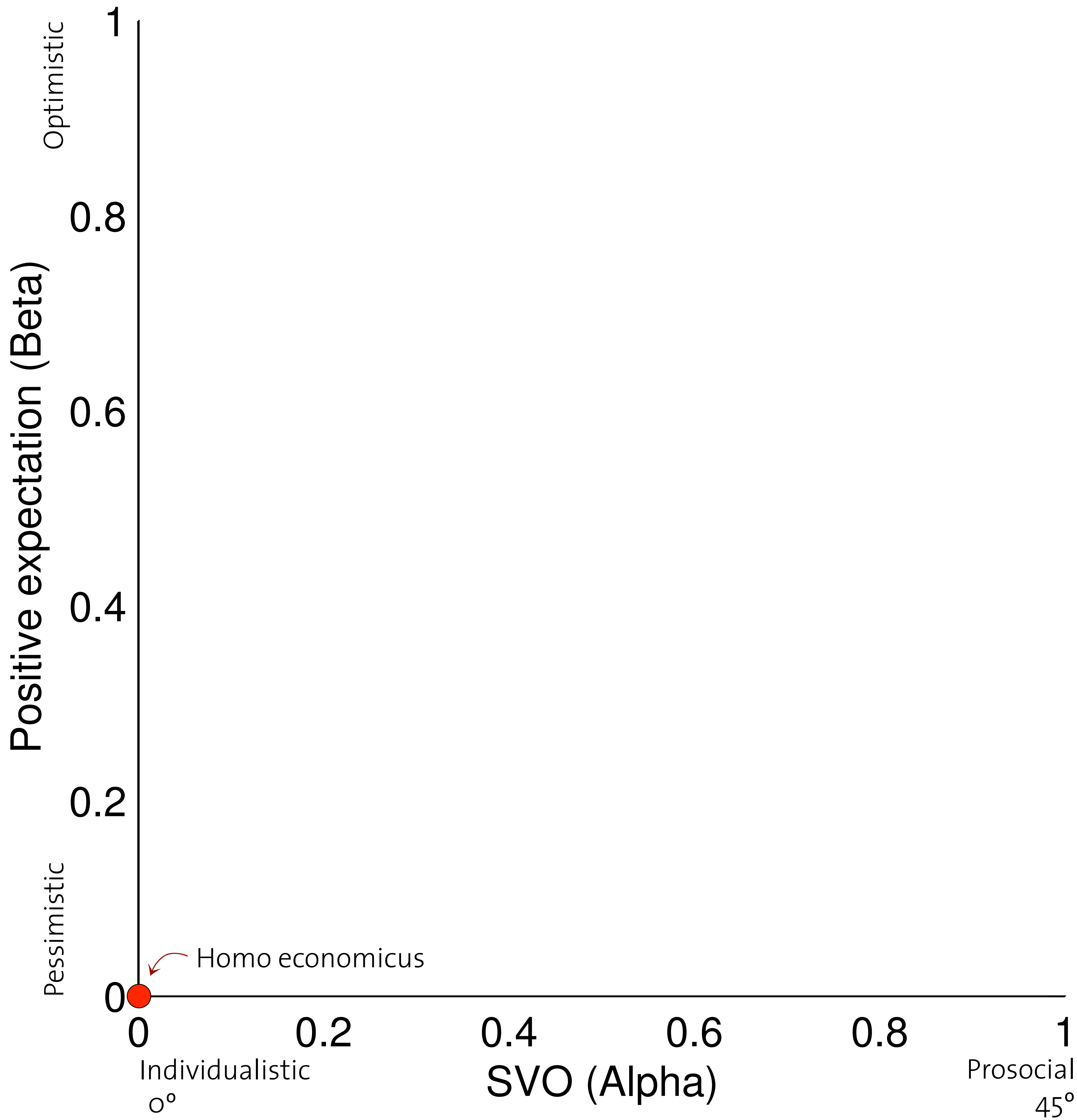
Bring this all together in a coherent and well integrated way

Build a framework that can rationalize trust based cooperation given heterogeneity in peoples' preferences and beliefs

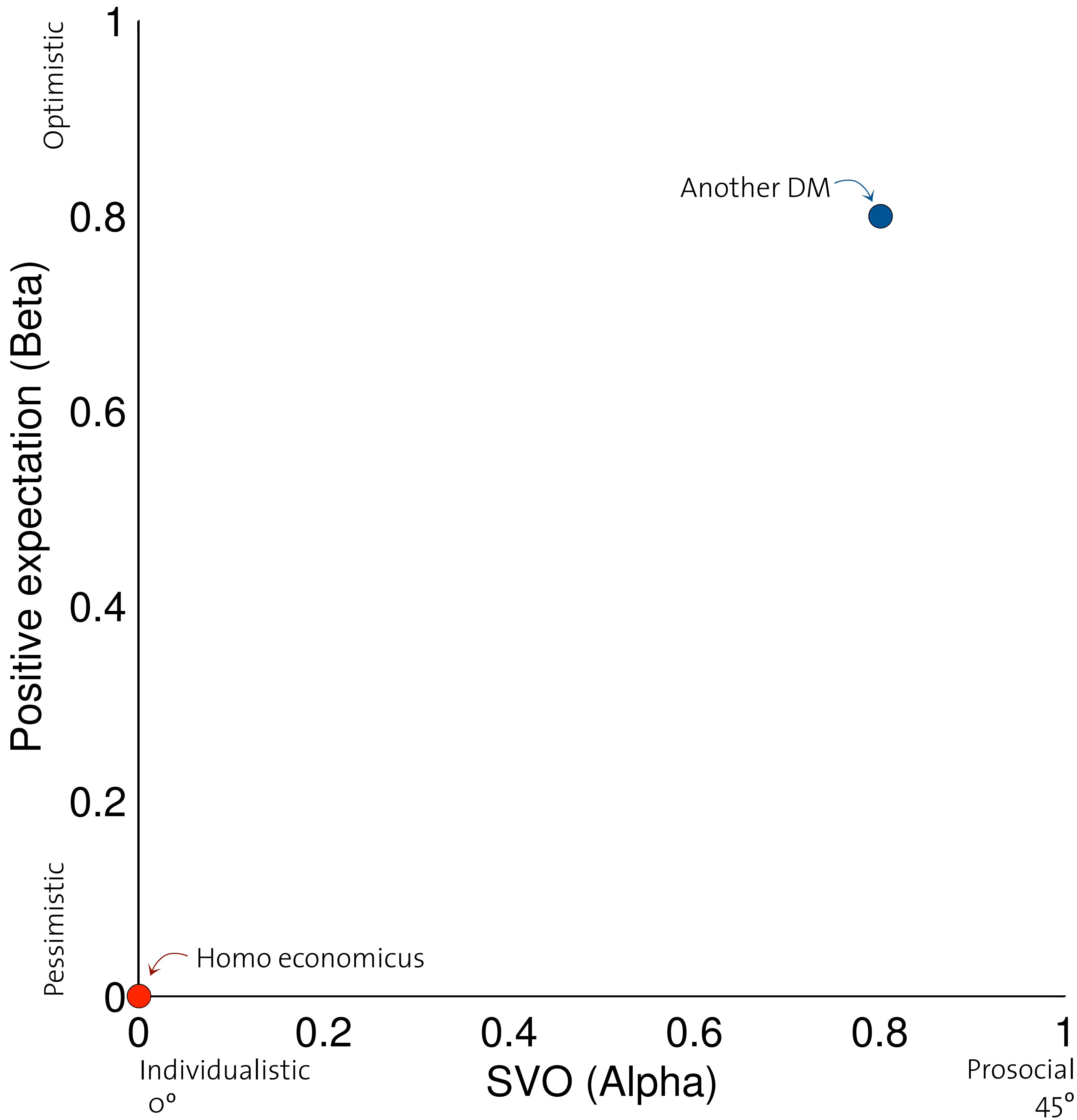
		Player 2	
		C	D
Player 1	C	.8, .8	0, 1
	D	1, 0	.6, .6



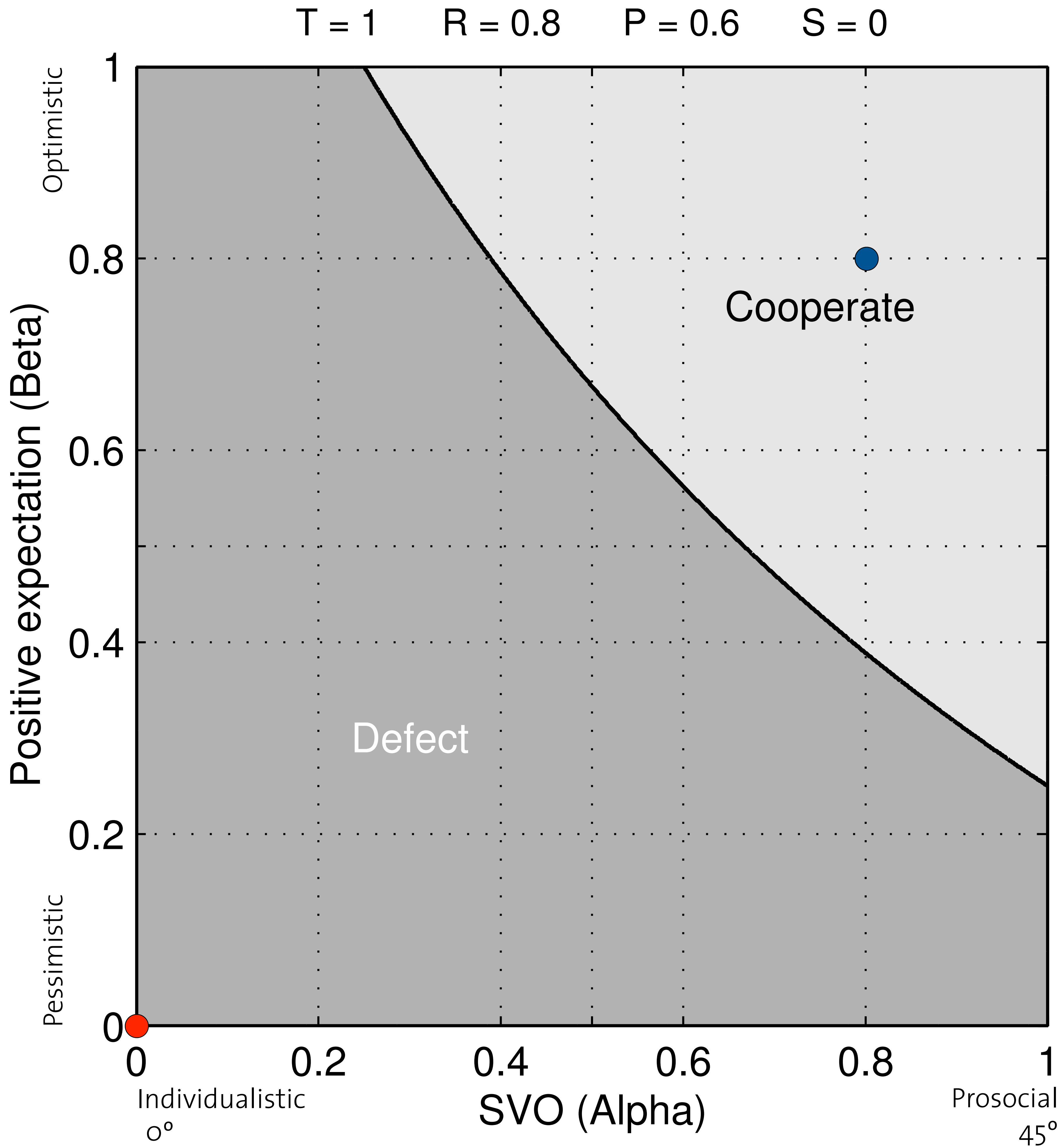
		Player 2	
		C	D
Player 1	C	.8, .8	0, 1
	D	1, 0	.6, .6



		Player 2	
		C	D
Player 1	C	.8, .8	0, 1
	D	1, 0	.6, .6



		Player 2	
		C	D
Player 1	C	.8, .8	0, 1
	D	1, 0	.6, .6



		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o$$

$$u(\mathrm{C}) = [\beta \cdot (R + \alpha \cdot R)] + [(1 - \beta) \cdot (S + \alpha \cdot T)]$$

$$u(\mathrm{D}) = [\beta \cdot (T + \alpha \cdot S)] + [(1 - \beta) \cdot (P + \alpha \cdot P)]$$

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o$$

Social preferences

$$u(\mathrm{C}) = [\beta \cdot (R + \alpha \cdot R)] + [(1 - \beta) \cdot (S + \alpha \cdot T)]$$

Trust

$$u(\mathrm{D}) = [\beta \cdot (T + \alpha \cdot S)] + [(1 - \beta) \cdot (P + \alpha \cdot P)]$$

Preferences and beliefs

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o$$

Social preferences

$$u(\mathrm{C}) = [\beta \cdot (R + \alpha \cdot R)] + [(1 - \beta) \cdot (S + \alpha \cdot T)]$$

$$=$$

$$u(\mathrm{D}) = [\beta \cdot (T + \alpha \cdot S)] + [(1 - \beta) \cdot (P + \alpha \cdot P)]$$

Preferences and beliefs

$$\Rightarrow$$

$$\beta_{crit} = \frac{P - S + \alpha P - \alpha T}{P + R - S - T + \alpha P + \alpha R - \alpha S - \alpha T}$$

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

$$u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o$$

Social preferences

$$u(\mathrm{C}) = [\beta \cdot (R + \alpha \cdot R)] + [(1 - \beta) \cdot (S + \alpha \cdot T)]$$

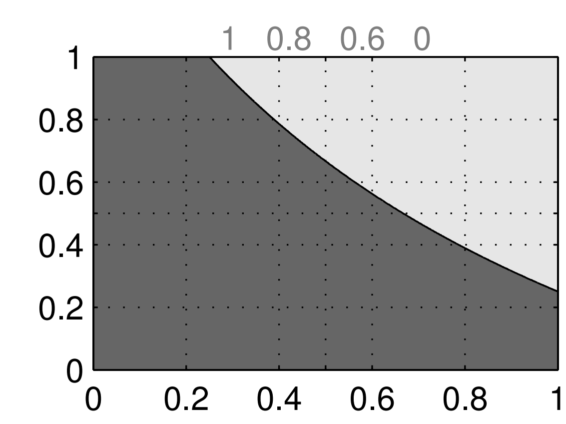
$$u(\mathrm{D}) = [\beta \cdot (T + \alpha \cdot S)] + [(1 - \beta) \cdot (P + \alpha \cdot P)]$$

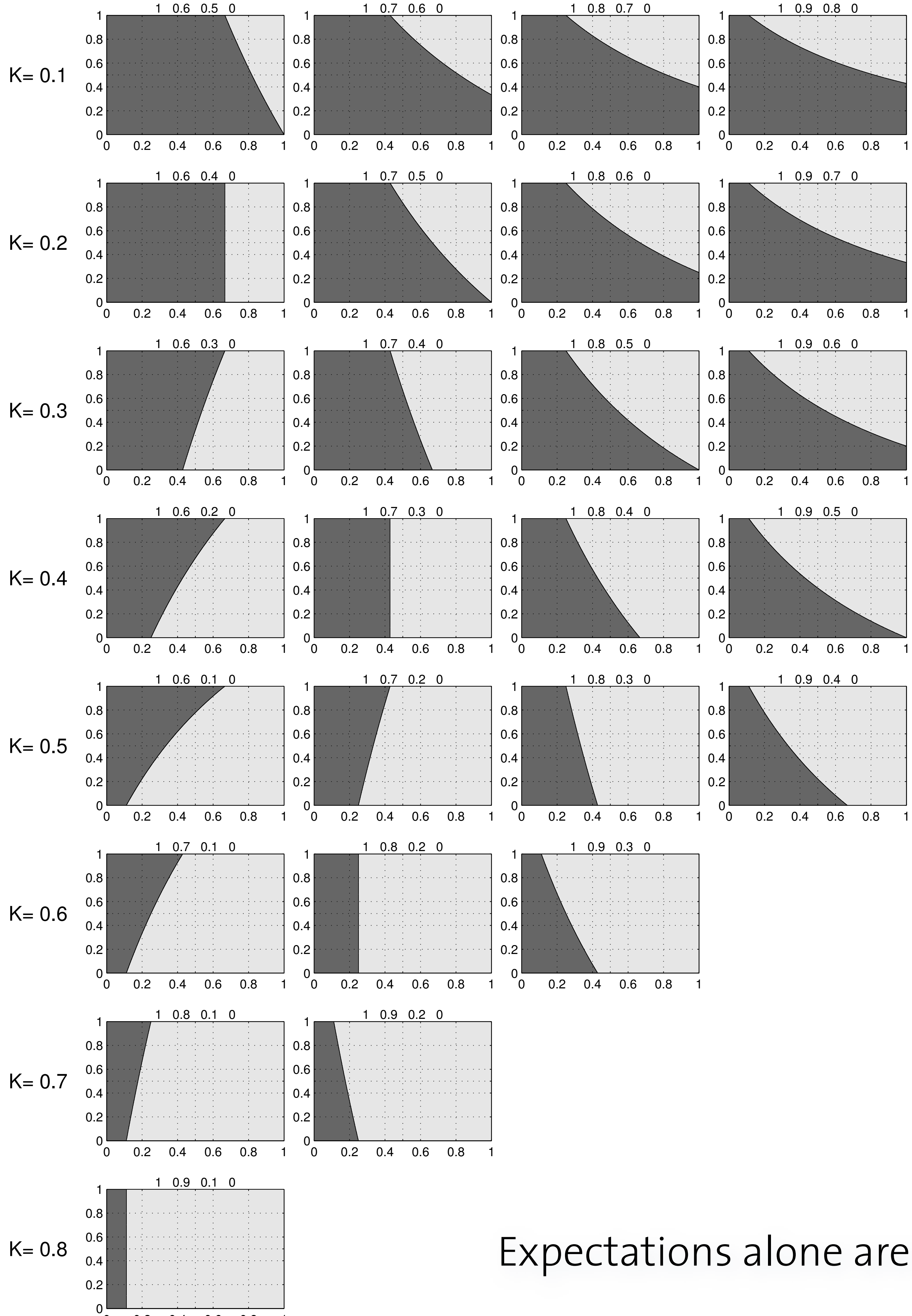
Preferences and beliefs

$$\beta_{crit} = \frac{P - S + \alpha P - \alpha T}{P + R - S - T + \alpha P + \alpha R - \alpha S - \alpha T}$$

Trust threshold

		Player 2	
		C	D
Player 1	C	.8, .8	0, 1
	D	1, 0	.6, .6





Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another.

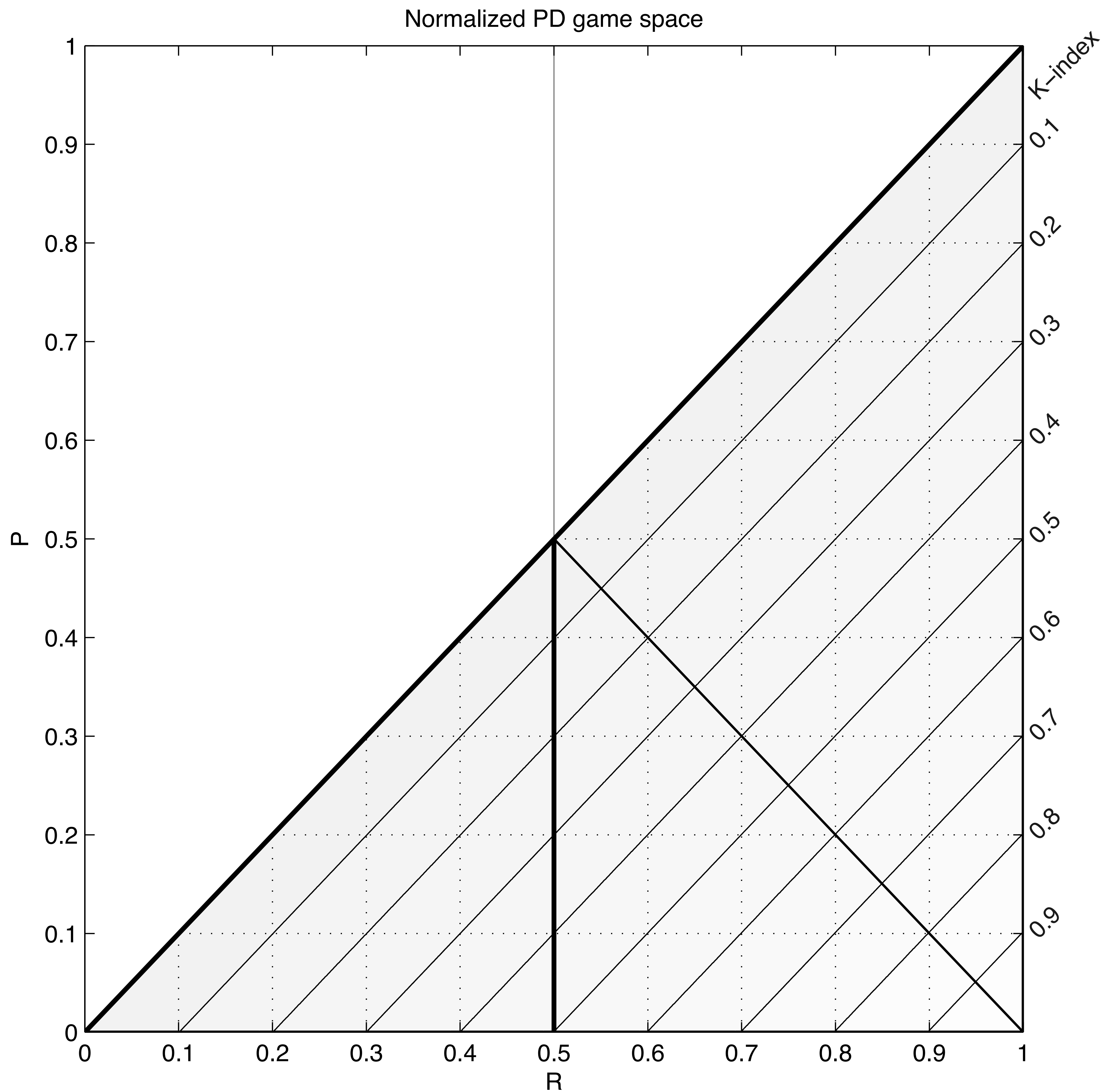
Rousseau et al., 1998

Expectations

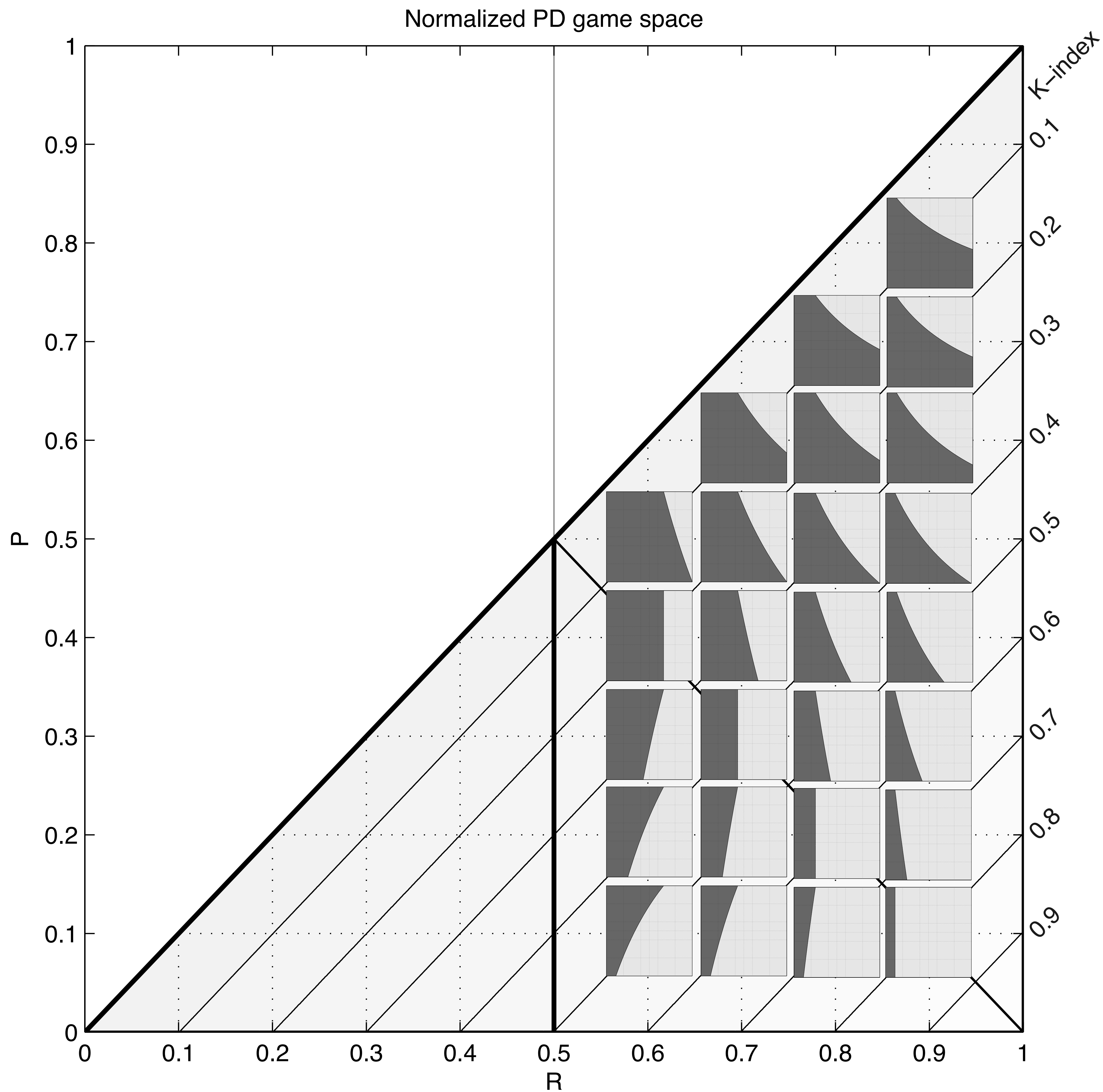
...with the intention of improving collective outcomes.

Preferences

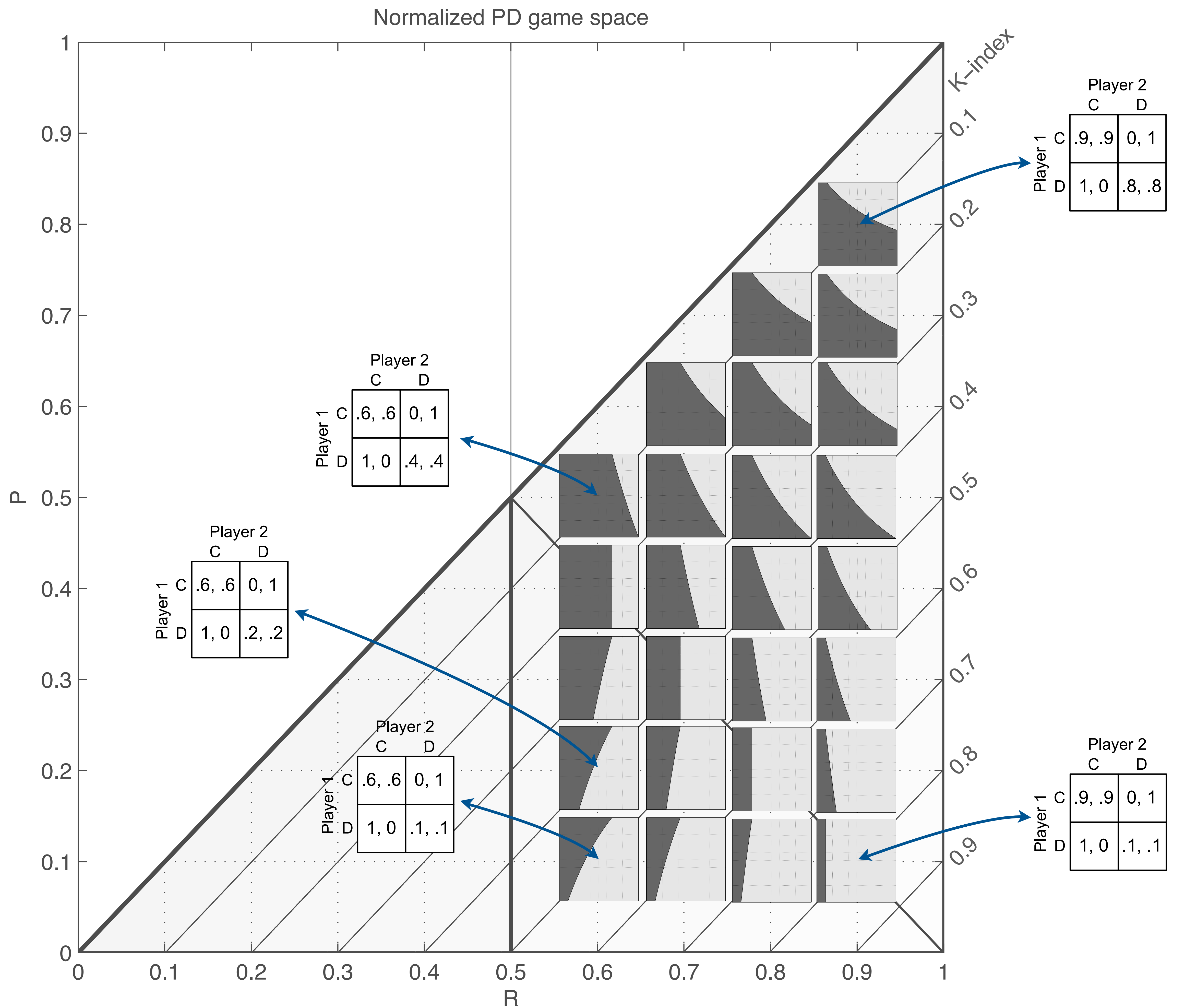
Expectations alone are insufficient to justify trust based cooperation
Social preferences are necessary also



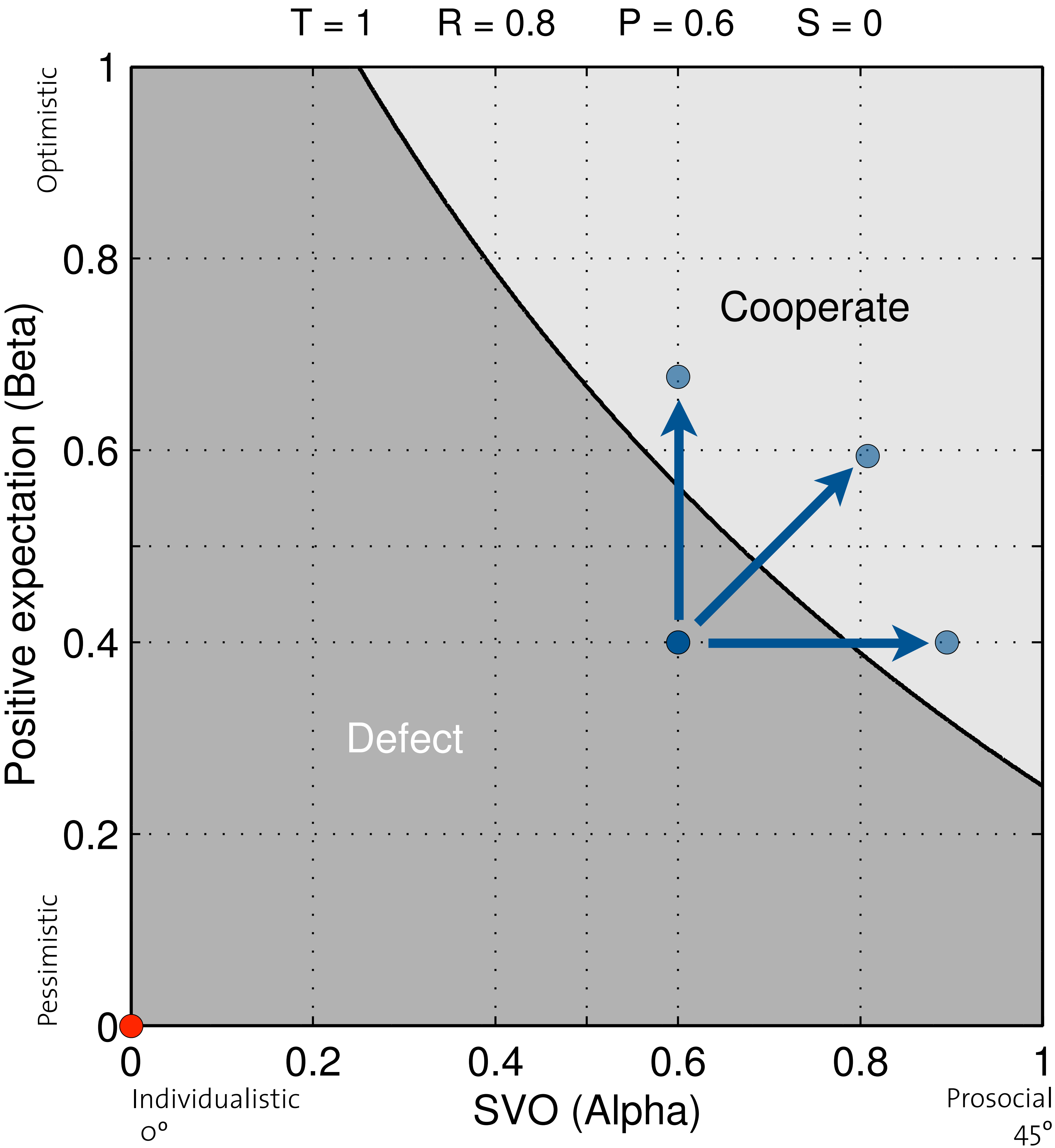
Normalized PD game space
K–index
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
P
R
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1



Normalized PD game space
P
R
K–index
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

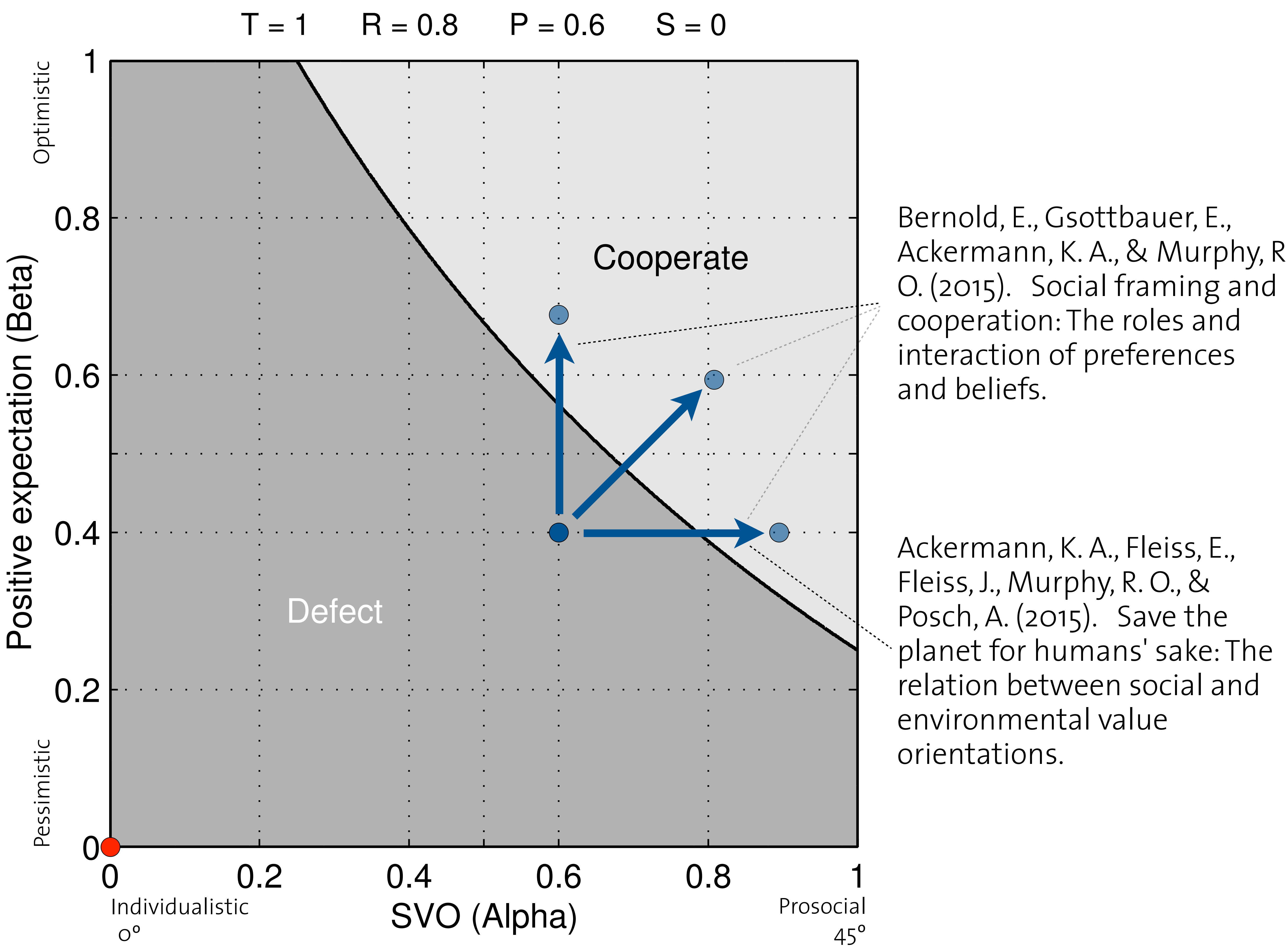


Normalized PD game space
K–index
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
P
R
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Player 2
C D
Player 1
C .9, .9 0, 1
D 1, 0 .8, .8
Player 2
C D
Player 1
C .6, .6 0, 1
D 1, 0 .4, .4
Player 2
C D
Player 1
C .6, .6 0, 1
D 1, 0 .2, .2
Player 2
C D
Player 1
C .6, .6 0, 1
D 1, 0 .1, .1
Player 2
C D
Player 1
C .9, .9 0, 1
D 1, 0 .1, .1



Mechanisms for nudging people to act more cooperatively

Interventions to change people's **preferences** and/or mechanisms to change their **beliefs**


T = 1
R = 0.8
P = 0.6
S = 0
Optimistic
Pessimistic
Positive expectation (Beta)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
Individualistic
0°
SVO (Alpha)
Prosocial
45°
Cooperate
Defect
Bernold, E., Gsottbauer, E., Ackermann, K. A., & Murphy, R. O. (2015). Social framing and cooperation: The roles and interaction of preferences and beliefs.
Ackermann, K. A., Fleiss, E., Fleiss, J., Murphy, R. O., & Posch, A. (2015). Save the planet for humans' sake: The relation between social and environmental value orientations.

Conclusions

- This started with a question:
  - How is SVO related to trust?
  - SVO is a necessary, but not sufficient, precursor to trust based cooperation.
  - If a DM is sufficiently prosocial, and has sufficient expectations that the other player will cooperate, then we expect trust based cooperation to emerge, provided the dilemma is not too severe.

$\alpha \quad \beta \quad TRPS$

Conclusions

- Develop a psychologically grounded model of trust based cooperation that is consistent with the principles of rationality and is measurable
- Integrate SVO (preferences), expectations (beliefs), trust based cooperation among interdependent players
- Derive precise trust thresholds over combinations of social preferences and beliefs
- Rapoport's K-index is isomorphic to the minimum level of SVO required by a DM to justify cooperation given a uniform prior (beliefs about the other player)
- Different PD games can have radically different properties

References

- Murphy, R. O., & Ackermann, K. A., (2015). Social preferences, positive expectations, and trust based cooperation. *Journal of Mathematical Psychology*. In press.

- Rapoport, A. (1967). A note on the "index of cooperation" for Prisoner's Dilemma. *Journal of Conflict Resolution*, 11(1), 100–103.

- Rousseau, D., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404.

- Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, 77(2), 337–349.

  - See also http://vlab.ethz.ch/svo for code and details about the measures and models.

PD game

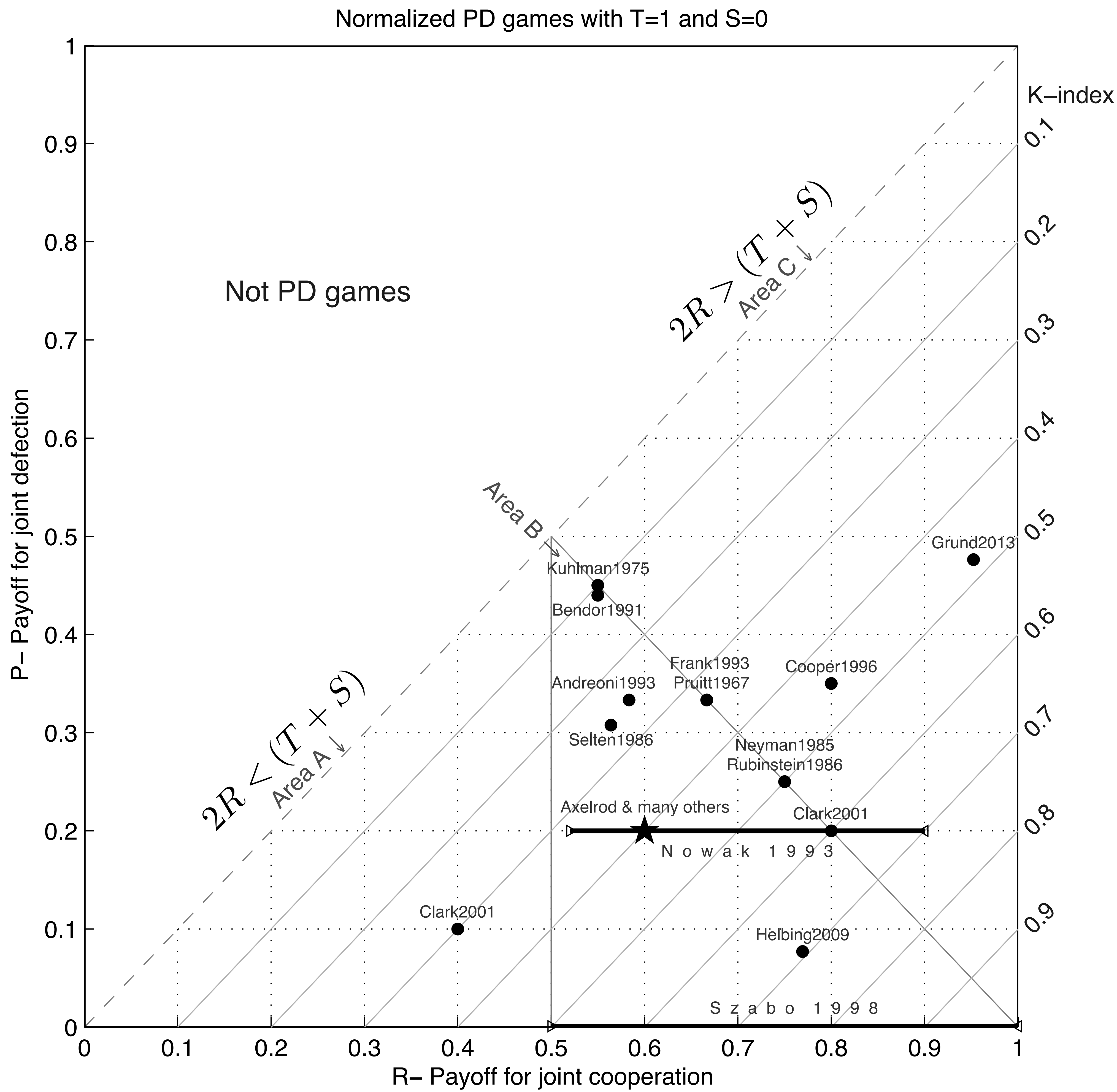
Player 2

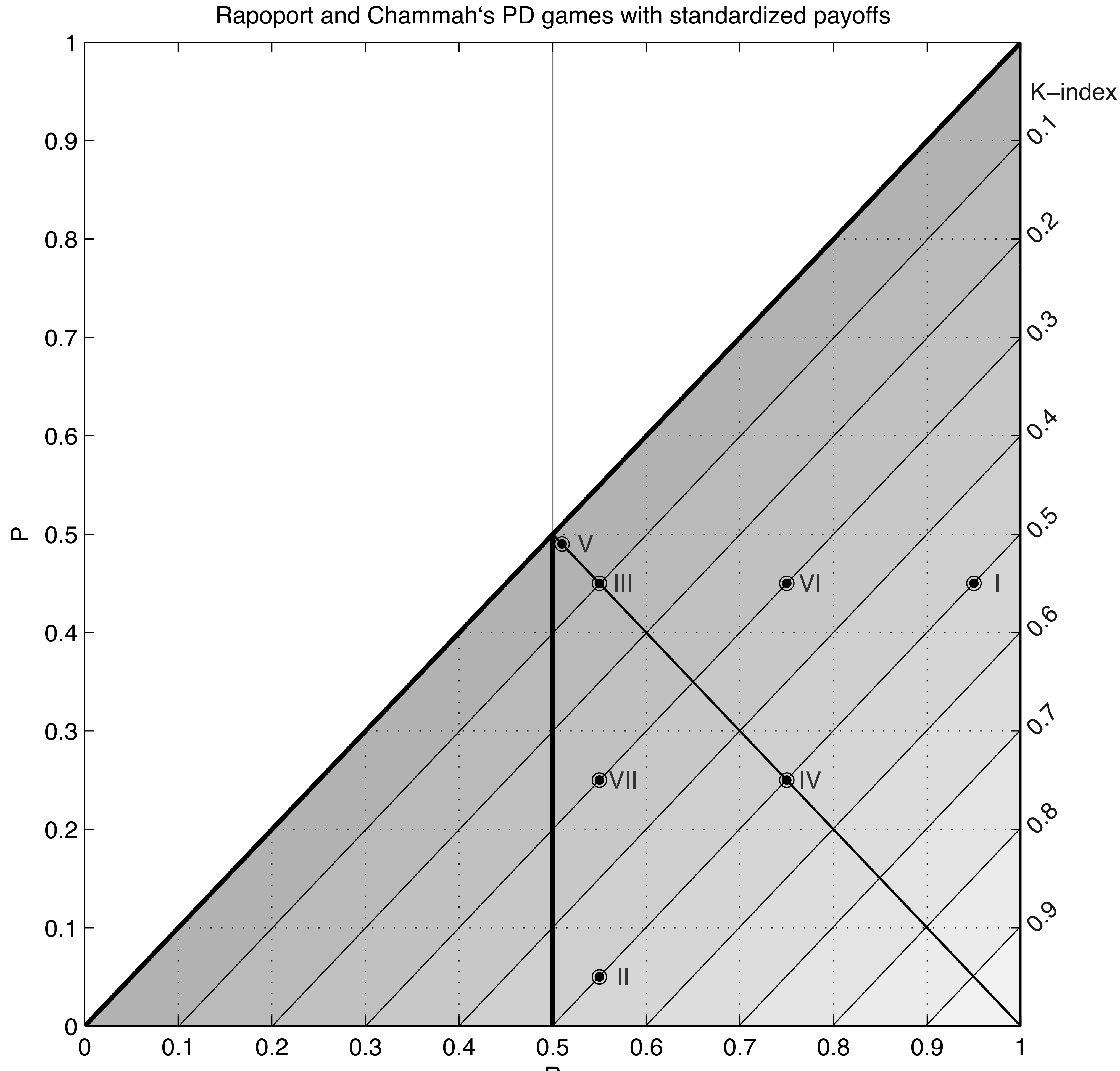
Player 1		C	D
	C	R, R	S, T
	D	T, S	P, P

$$T > R > P > S$$

Normalizing the payoffs to be between 0 and 1

$$\left[T_n = \frac{T-S}{T-S} = 1, \quad R_n = \frac{R-S}{T-S}, \quad P_n = \frac{P-S}{T-S}, \quad S_n = \frac{S-S}{T-S} = 0\right]$$


Normalized PD games with T=1 and S=0
K−index
Not PD games
2R > (T + S)
Area C ↓
2R < (T + S)
Area A ↓
Area B →
Kuhlman1975
Bendor1991
Grund2013
Frank1993
Pruitt1967
Cooper1996
Andreoni1993
Selten1986
Neyman1985
Rubinstein1986
Axelrod & many others
Clark2001
Nowak 1993
Clark2001
Helbing2009
Szabo 1998
P− Payoff for joint defection
R− Payoff for joint cooperation



Rapoport and Chammah's PD games with standardized payoffs
K–index
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
P
R
V
III
VI
I
VII
IV
II

- Rapoport's K-index is isomorphic of the minimum level of SVO required by a DM to justify cooperation given a uniform prior (beliefs about the other player)
- For any PD game, assuming a uniform prior, the critical level of SVO can be found with:

$$\alpha_{crit} = -\frac{P - R - S + T}{P - R + S - T}$$

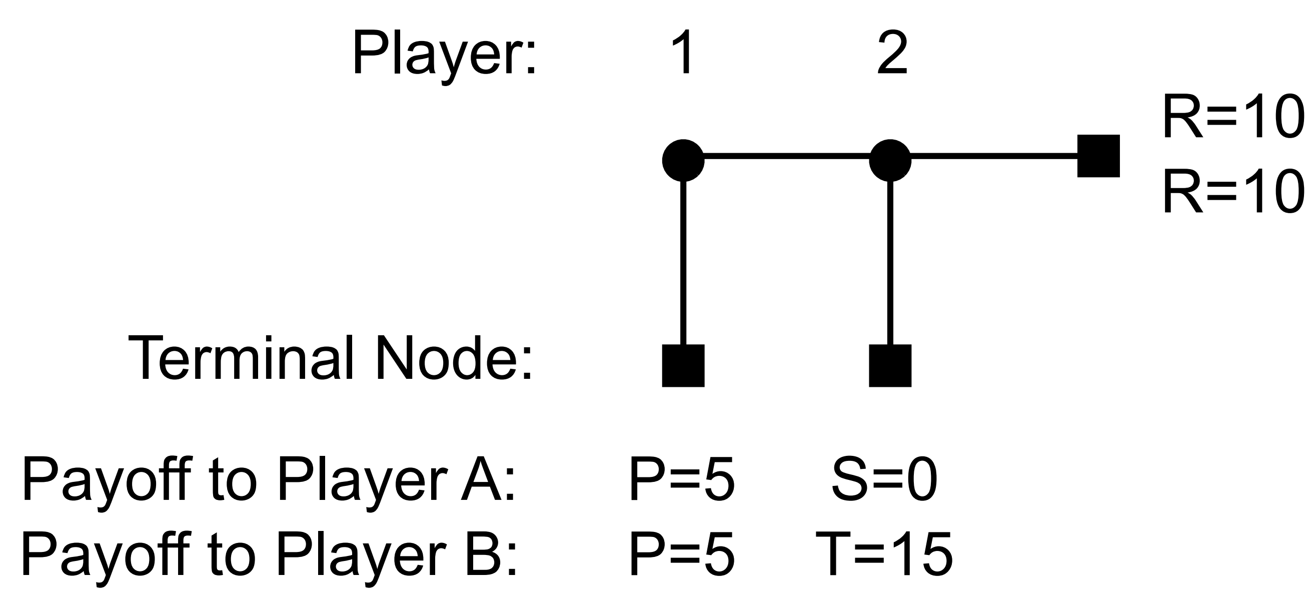
Trust

Player 2

Player 1	C	D
C	R, R	S, T
D	T, S	P, P

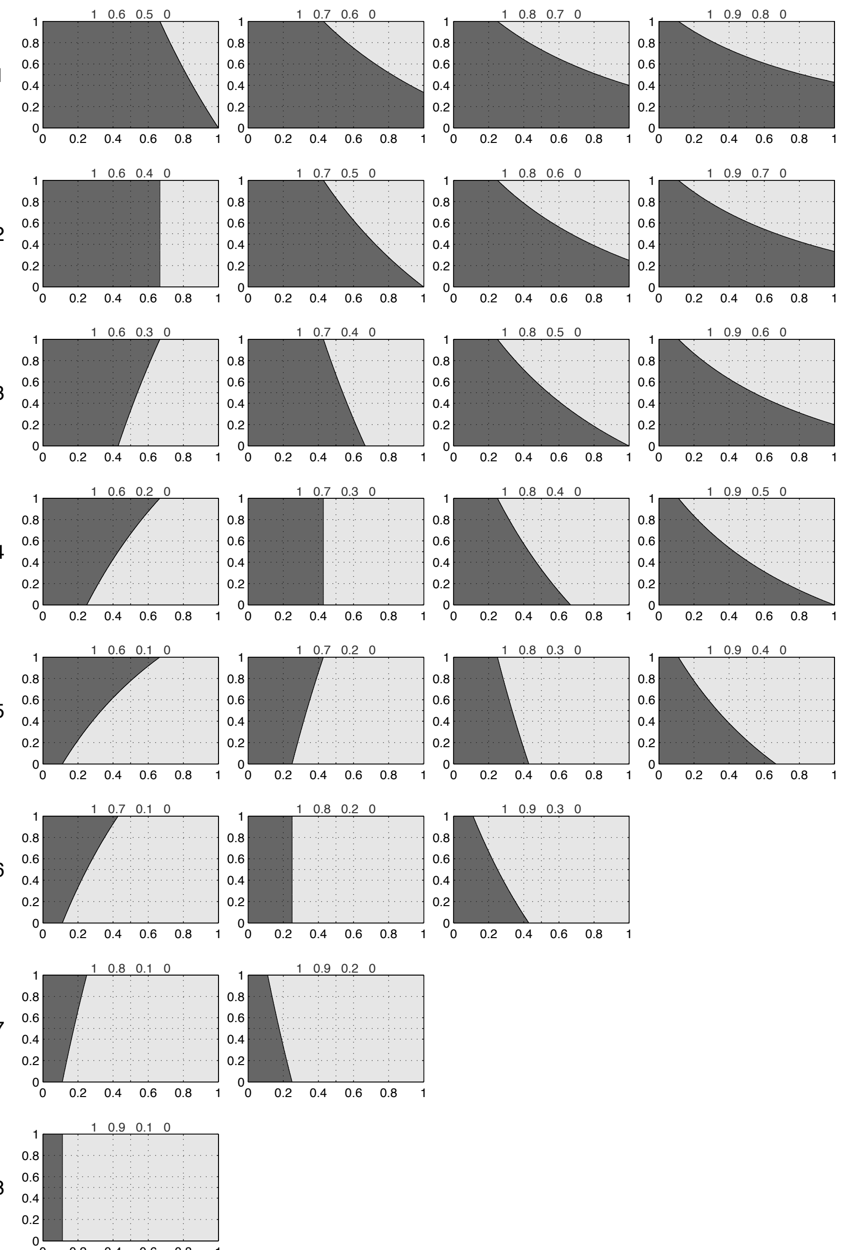
$$T > R > P > S$$

$$2R > (T + S)$$



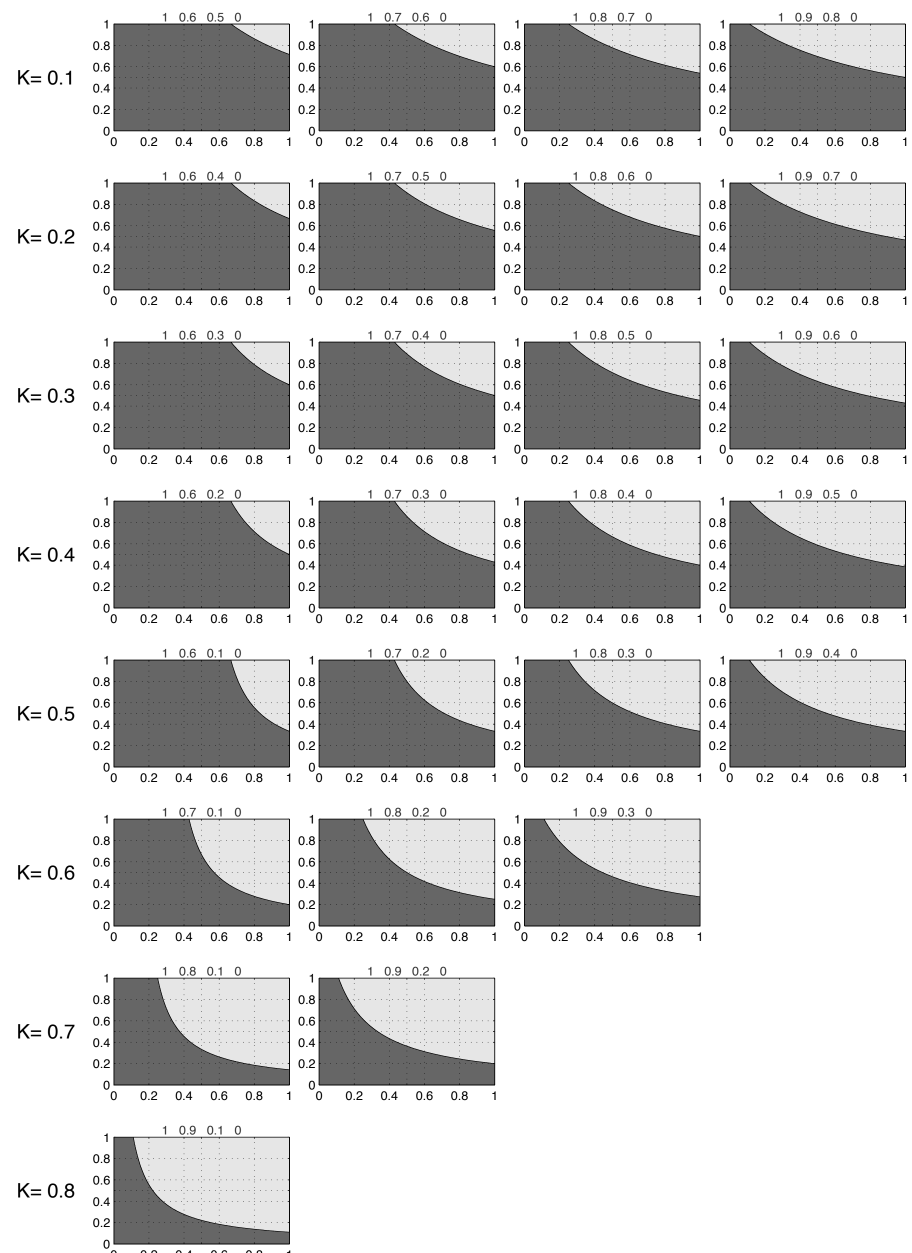
$$\pi_s + \alpha \cdot \pi_o$$

Simple joint utility function



$$\pi_s + \mathbb{1}\alpha\pi_o$$

Contigent function



$$(\pi_s^{1-\alpha}) \cdot (\pi_o^{\alpha})$$

Cobb–Douglas function