Toulouse School of Economics

Evolution of Economic Behavior

TSE M1 – Semester 1

September 2023

Paul Seabright

Week 4: The Cognitive and Emotional Foundations of Cooperation.

Economics for the Common Good

Recapping: the explanation of declining violence from behavioral economics and neuroscience

- Reason has not *replaced* emotion but has *harnessed* it
- Purely cognitive approaches to the enforcement of trust cannot work (the reliability of the reprisal mechanisms depends on emotional components)
- In particular, trust is more effective in the presence of strong reciprocity
- But effectively designed institutions can make a little reciprocity go a long way

What are the mechanisms?

- Until recently the emphasis has been on cultural factors, as in the Elias story.
- However, Wrangham (*The Goodness Paradox*, 2019) has revived the debate over genetic factors, notably through the “domestication” hypothesis.
- To understand this, note that humans compare very differently to other species on “reactive” versus “proactive” violence: we score low on reactive and high on proactive violence compared to (for example) chimpanzees.
- He suggests that by deliberately executing men who were unable to control their reactive violence, human societies domesticated us, with symptoms familiar from the domestication of other species lives wolves and foxes.
- Russian geneticist Dmitry Belyaev had shown in an experiment beginning in 1959 that selection of wild foxes for friendliness to humans produced tame, dog-like characteristics (including white patches and floppy ears) up to 70% in 30 generations.

The silver fox experiment of Dmitry Belyaev

Wild silver fox

Domesticated silver fox

The cognitive and emotional foundations of cooperation (I): Outline

- What's needed for trust
- The key to our psychological trade-offs
- A view from behavioral economics
- Supporting evidence from neuroscience
- Conclusions: how is this evidence consistent with natural selection?

The strange case of Phineas Gage:

- Phineas Gage (1823-1860) was an American railroad construction foreman who survived a freak accident in Vermont that drove a large iron spike through his skull, destroying most of the left frontal lobe of his brain.

- He recovered completely physically but was reported afterwards to have become psychologically changed. Physician John Harlow, who knew him before and after, wrote:

The equilibrium or balance, so to speak, between his intellectual faculties and animal propensities, seems to have been destroyed. He is fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom), manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires, at times pertinaciously obstinate, yet capricious and vacillating, devising many plans of future operations, which are no sooner arranged than they are abandoned in turn for others appearing more feasible. A child in his intellectual capacity and manifestations, he has the animal passions of a strong man. Previous to his injury, although untrained in the schools, he possessed a well-balanced mind, and was looked upon by those who knew him as a shrewd, smart business man, very energetic and persistent in executing all his plans of operation. In this regard his mind was radically changed, so decidedly that his friends and acquaintances said he was "no longer Gage.”

Gage and his accident with the tamping iron:

Controversy surrounding the interpretation of Gage's condition:

- Antonio Damasio in *Descartes' Error: Emotion, Reason and the Human Brain* (1994) cites Gage in support of his *somatic marker hypothesis*.

- This holds that feelings in the body are associated with emotions, are located in the ventromedial prefrontal cortex, and evolved to guide and give coherence to decision making.

- Damasio suggested Gage was similar to patients with frontal lobe damage who have unimpaired intellect but (according to him) are incapable of feeling and therefore of coherent decision-making.

- This portrayal of Gage has been strongly criticized, for example by Malcolm Macmillan in *An Odd Kind of Fame: Stories of Phineas Gage* (2000) as selective and distorted to fit the hypothesis.

- The scant evidence makes it hard to judge Gage's case, and the issue of the role of emotions in decision-making remains complex.

A role for the emotions: they're needed for trust:

- It's not enough to be good at spotting who can be trusted
- We also have to be good at inspiring trust in others
- High cognitive skills do not necessarily help us do this
- Kaushik Basu and the taxi driver
- Our solution:
  - An evolved cognitive AND emotional psychology
  - Trust in *institutions*

The key to our psychological trade-offs

- Cognitive capacities are exquisitely context-sensitive but no good for making commitment

- Recent evidence from experimental psychology and neurophysiology suggests *emotion* plays an important role in social cooperation, which was vital to our ancestors' survival

- It also suggests that many of the skills that promote cooperation are adapted modules of our brain, not just forms of general-purpose rationality

- Like chimps, we avoid violence when it doesn't pay – but we have more elaborate mechanisms to stop it from paying

A view from behavioral economics

- Cooperation needs discrimination PLUS commitment
- Three robust results from experimental behavioral economics:
- 1) Many (but not all) subjects are generous to strangers
- 2) Many (but not all) subjects display strong reciprocity
- 3) In repeated public goods games, cooperation starts positively but declines over time as subjects react negatively to others' free-riding – unless free-riders can be punished, even at a cost to the punishers!

From Henrich et al, “In Search of Homo Economics: Behavioral Experiments in 15 small-scale societies”, American Economic Review 2001

TABLE 1—THE ULTIMATUM GAME IN SMALL-SCALE SOCIETIES

Group	Country	Mean offer[a]	Modes[b]	Rejection rate[c]	Low-offer rejection rate[d]
Machiguenga	Peru	0.26	0.15/0.25 (72)	0.048 (1/21)	0.10 (1/10)
Hadza (big camp)	Tanzania	0.40	0.50 (28)	0.19 (5/26)	0.80 (4/5)
Hadza (small camp)	Tanzania	0.27 (38)	0.20 (8/29)	0.28 (5/16)	0.31
Tsimané	Bolivia	0.37	0.5/0.3/0.25 (65)	0.00 (0/70)	0.00 (0/5)
Quichua	Ecuador	0.27	0.25 (47)	0.15 (2/13)	0.50 (1/2)
Torguud	Mongolia	0.35	0.25 (30)	0.05 (1/20)	0.00 (0/1)
Khazax	Mongolia	0.36	0.25		
Mapuche	Chile	0.34	0.50/0.33 (46)	0.067 (2/30)	0.2 (2/10)
Au	PNG	0.43	0.3 (33)	0.27 (8/30)	1.00 (1/1)
Gnau	PNG	0.38	0.4 (32)	0.4 (10/25)	0.50 (3/6)
Sangu farmers	Tanzania	0.41	0.50 (35)	0.25 (5/20)	1.00 (1/1)
Sangu herders	Tanzania	0.42	0.50 (40)	0.05 (1/20)	1.00 (1/1)
Unresettled villagers	Zimbabwe	0.41	0.50 (56)	0.1 (3/31)	0.33 (2/5)
Resettled villagers	Zimbabwe	0.45	0.50 (70)	0.07 (12/86)	0.57 (4/7)
Achuar	Ecuador	0.42	0.50 (36)	0.00 (0/16)	0.00 (0/1)
Orma	Kenya	0.44	0.50 (54)	0.04 (2/56)	0.00 (0/0)
Aché	Paraguay	0.51	0.50/0.40 (75)	0.00 (0/51)	0.00 (0/8)
Lamelara[e]	Indonesia	0.58	0.50 (63)	0.00 (3/8)	0.00 (4/20)

Note: PNG = Papua New Guinea.

[a] This column shows the mean offer (as a proportion) in the ultimatum game for each society.

[b] This column shows the modal offer(s), with the percentage of subjects who make modal offers (in parentheses).

[c] The rejection rate (as a proportion), with the actual numbers given in parentheses.

[d] The rejection rate for offers of 20 percent or less, with the actual numbers given in parentheses.

[e] Includes experimenter-generated low offers.

Machiguenga	Peru	0.26	0.15/0.25 (72)	0.048 (1/21)	0.10 (1/10)
Hadza (big camp)	Tanzania	0.40	0.50 (28)	0.19 (5/26)	0.80 (4/5)
Hadza (small camp)	Tanzania	0.27 (38)	0.20 (8/29)	0.28 (5/16)	0.31
Tsimané	Bolivia	0.37	0.5/0.3/0.25 (65)	0.00 (0/70)	0.00 (0/5)
Quichua	Ecuador	0.27	0.25 (47)	0.15 (2/13)	0.50 (1/2)
Torguud	Mongolia	0.35	0.25 (30)	0.05 (1/20)	0.00 (0/1)

From Fehr & Gaechter, “Cooperation and Punishment in Public Goods Experiments”, American Economic Review 2000

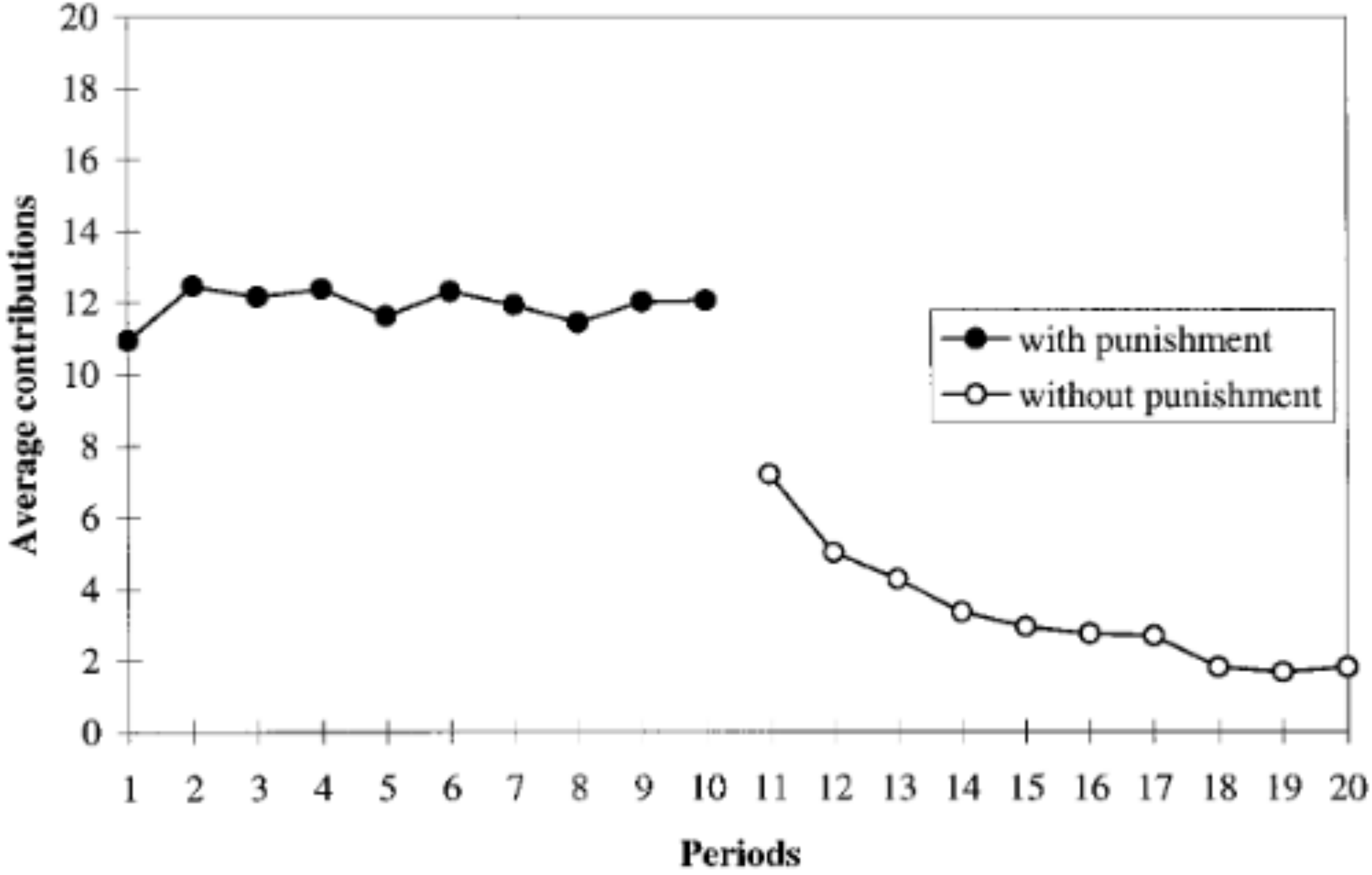


FIGURE 1A. AVERAGE CONTRIBUTIONS OVER TIME IN THE STRANGER-TREATMENT (SESSIONS 1 AND 2)

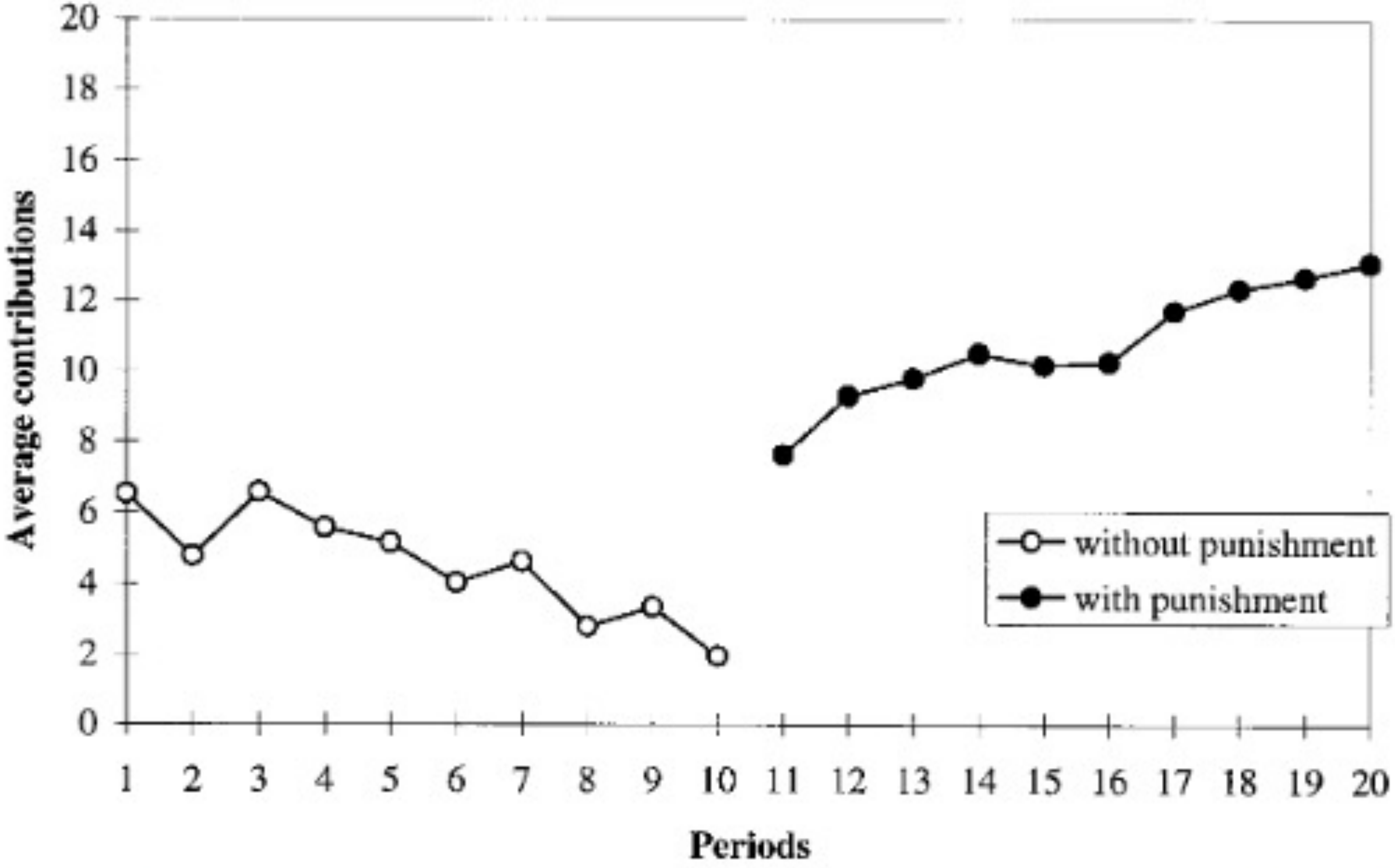


FIGURE 1B. AVERAGE CONTRIBUTIONS OVER TIME IN THE STRANGER-TREATMENT (SESSION 3)

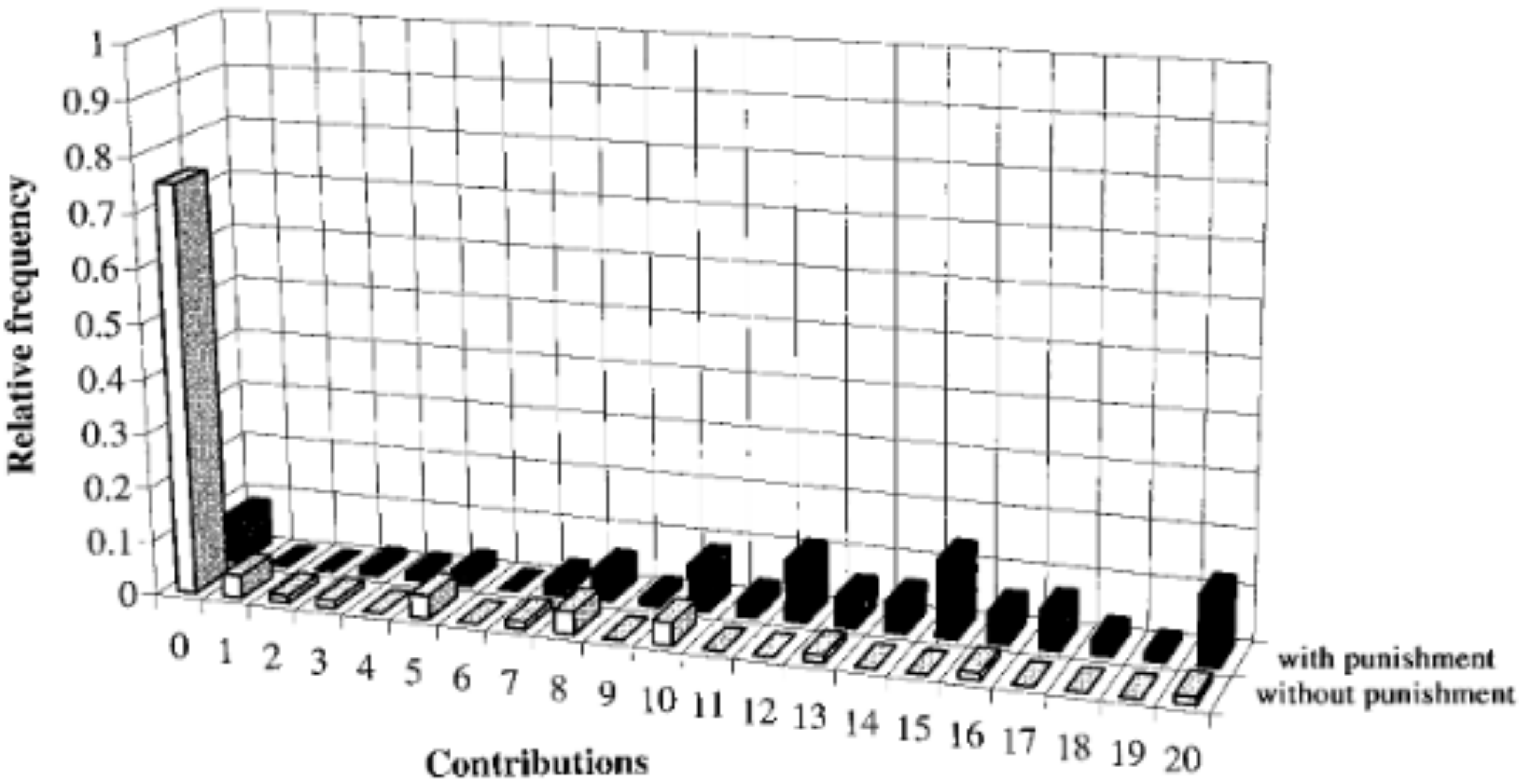


FIGURE 2. DISTRIBUTION OF CONTRIBUTIONS IN THE FINAL PERIODS OF THE STRANGER-TREATMENT WITH AND WITHOUT PUNISHMENT

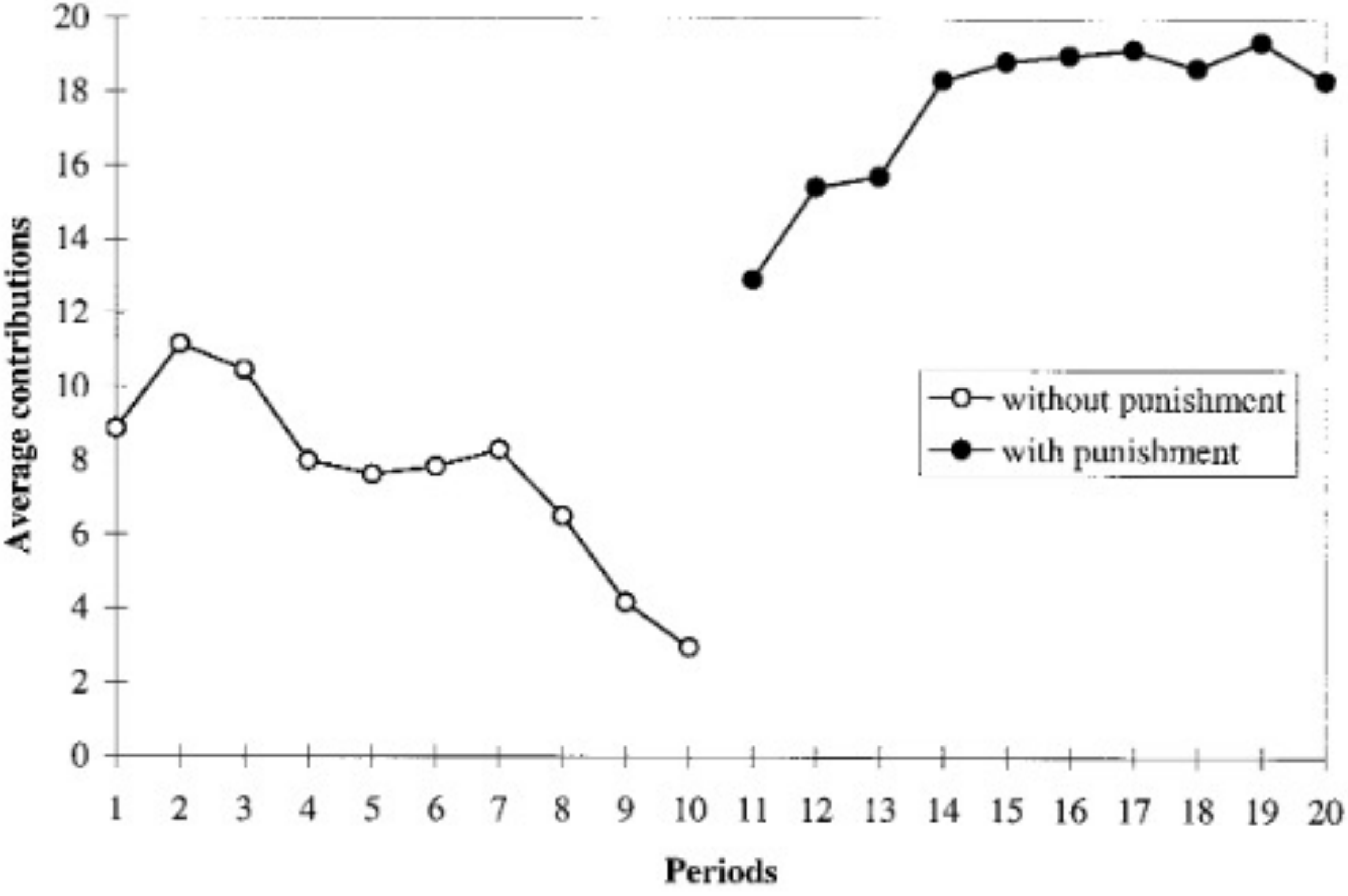


FIGURE 3B. AVERAGE CONTRIBUTIONS OVER TIME IN THE PARTNER-TREATMENT (SESSION 5)

Supporting evidence from neuroscience

- Commitment needs a neural mechanism
- Brain tissue is expensive, so our ancestors needed economical ways of encoding such behavior, either in cognitive short-cuts (for cheater detection) or in emotions (for commitment)
- Natural selection has repeatedly recruited existing neural machinery (eg homeostatic mechanisms) for strategic purposes (see Churchland: *Brain Trust*, Princeton 2011)
- Neuroscientific evidence is accumulating that commitment is linked with reward circuits in the brain

Anatomical separation of exploratory and exploitative decisions in the brain (Source: Dow et.al., Nature, June 15 2006)

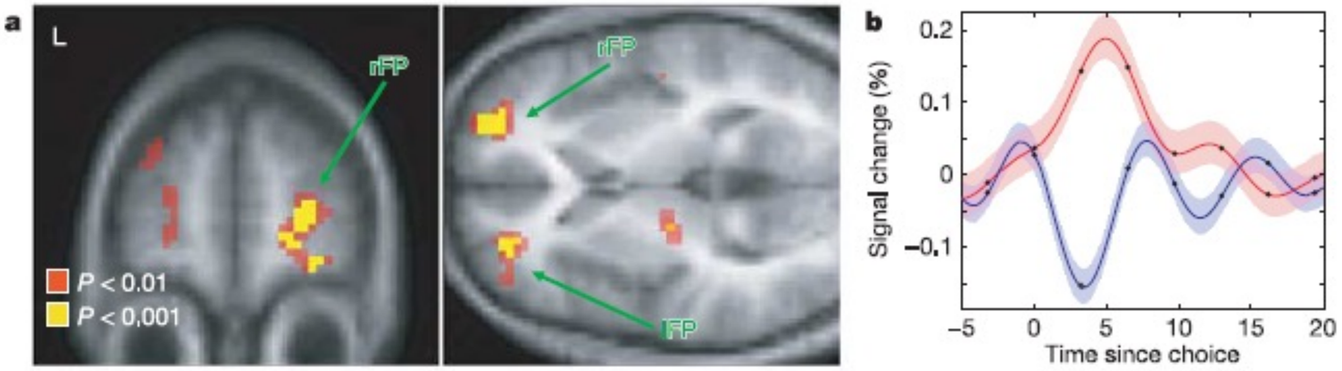


Figure 3 | Exploration-related activity in frontopolar cortex. **a**, Regions of left and right frontopolar cortex (lFP, rFP) showing significantly increased activation on exploratory compared with exploitative trials. Activation maps (yellow, $P < 0.001$; red, $P < 0.01$) are superimposed on a subject-averaged structural scan. The coordinates of activated areas are [−27,48,4, peak $z = 3.49$] for lFP and [27,57,6, peak $z = 4.13$] for rFP. **b**, rFP BOLD time courses averaged over 1,515 exploratory (red line) and 2,646 exploitative (blue line) decisions. Black dots indicate the sampling frequency (although, because sample alignment varied from trial to trial, time courses were upsampled). Coloured fringes show error bars (representing s.e.m.).

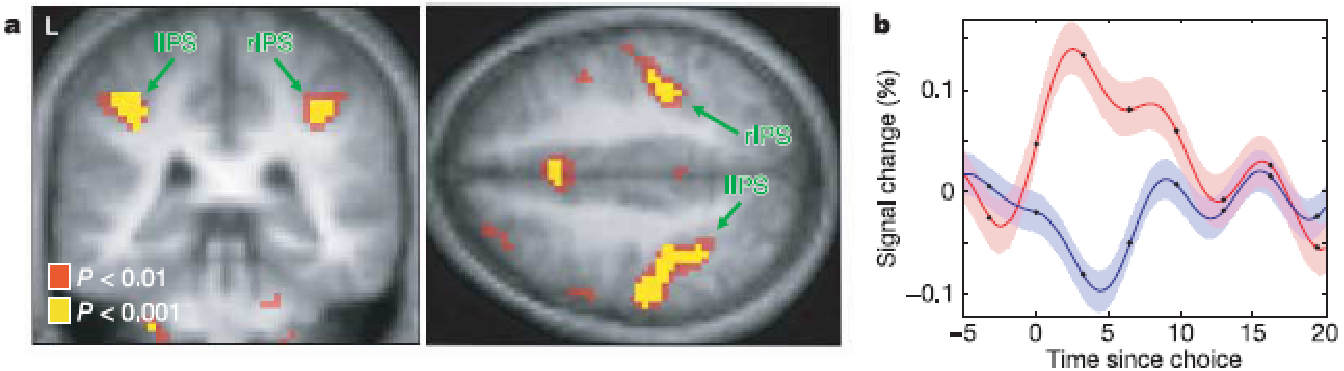
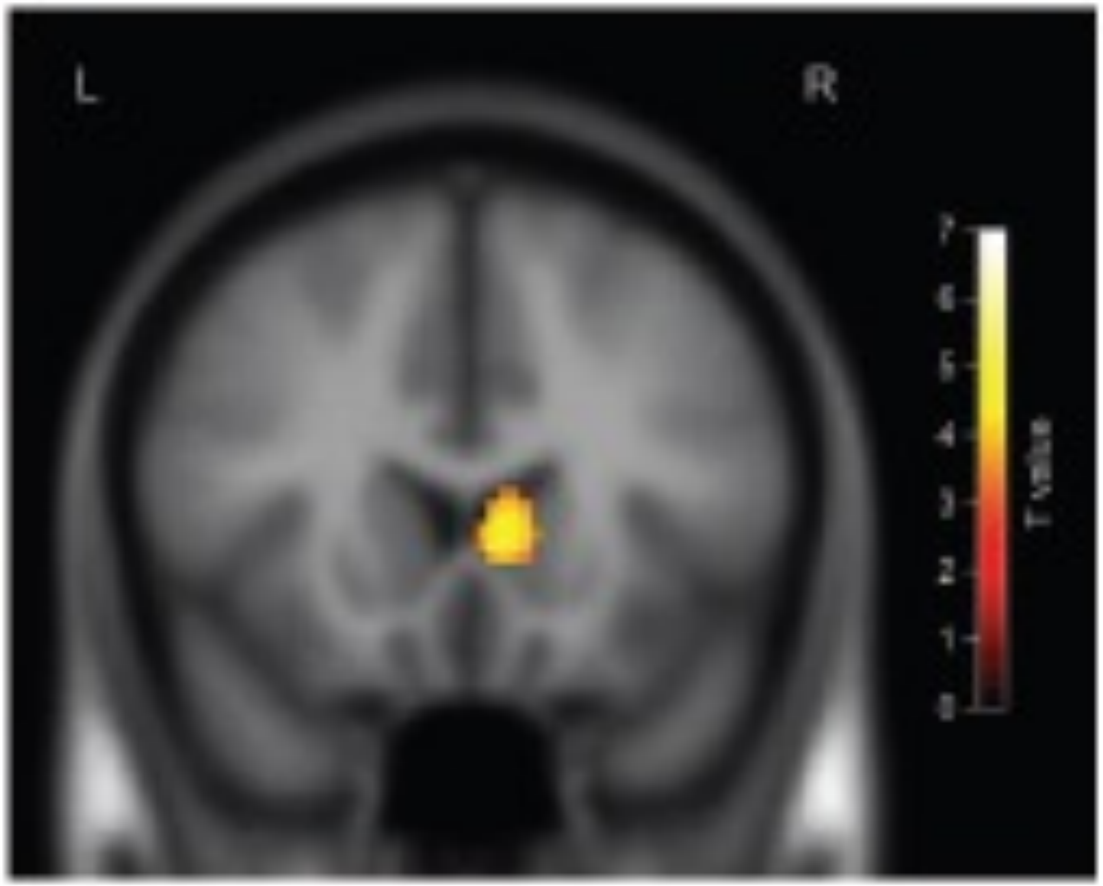


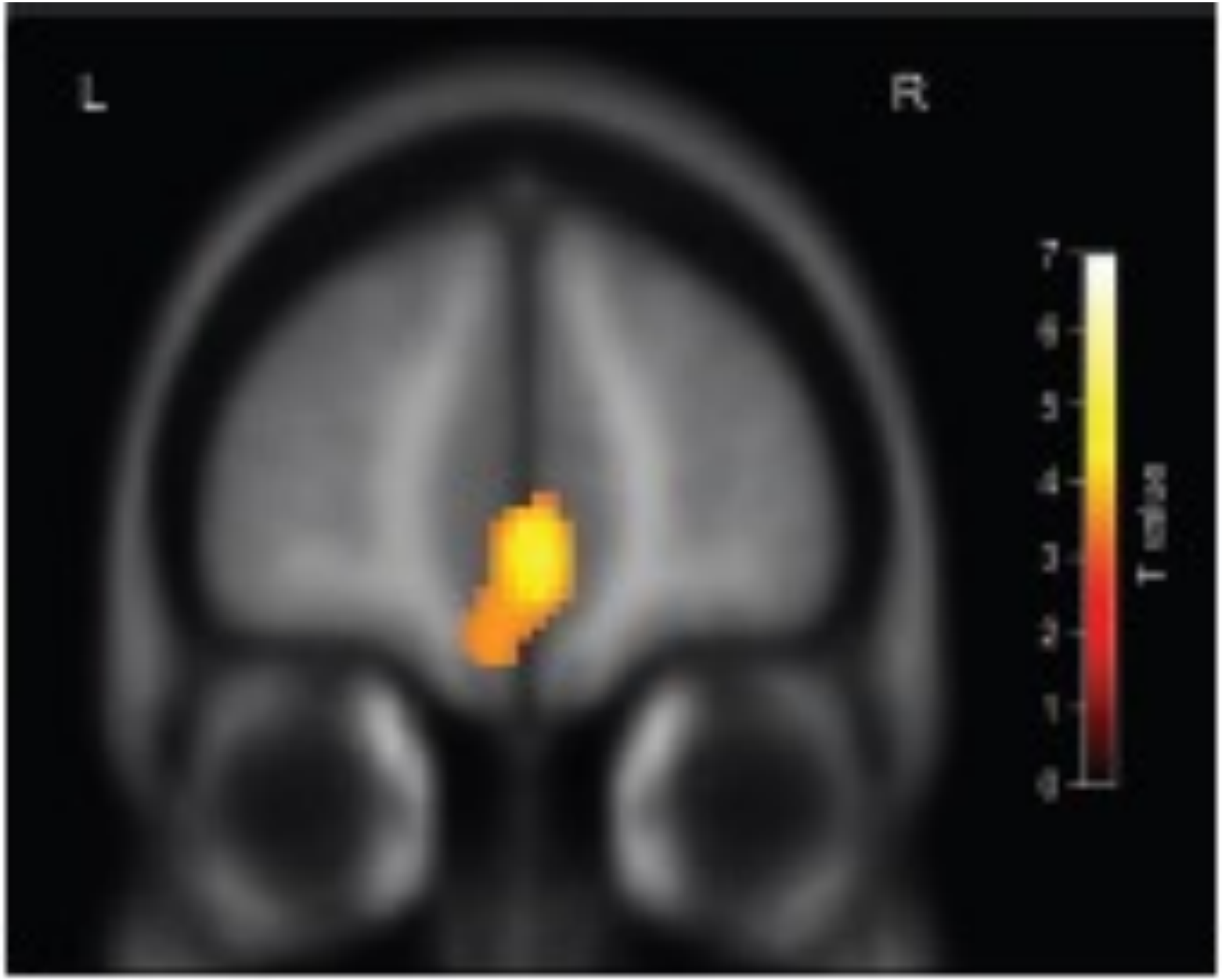
Figure 4 | Exploration-related activity in intraparietal sulcus. **a**, Regions of left and right intraparietal sulcus (lIPS and rIPS) showing significantly increased activation on exploratory compared with exploitative trials. Activation maps (yellow, $P < 0.001$; red, $P < 0.01$) are superimposed on a subject-averaged structural scan. The coordinates of the activated areas are [−29,−33,45, peak $z = 4.39$] for lIPS and [39,−36,42, peak $z = 4.16$] for rIPS. **b**, lIPS BOLD time courses averaged over 1,515 exploratory (red line) and 2,646 exploitative (blue line) decisions. Black dots indicate the sampling frequency (although, because sample alignment varied from trial to trial, time courses were upsampled). Coloured fringes show error bars (representing s.e.m.).

The neural basis of altruistic punishment
(Source: de Quervain et.al., *Science*, August 27 2004)

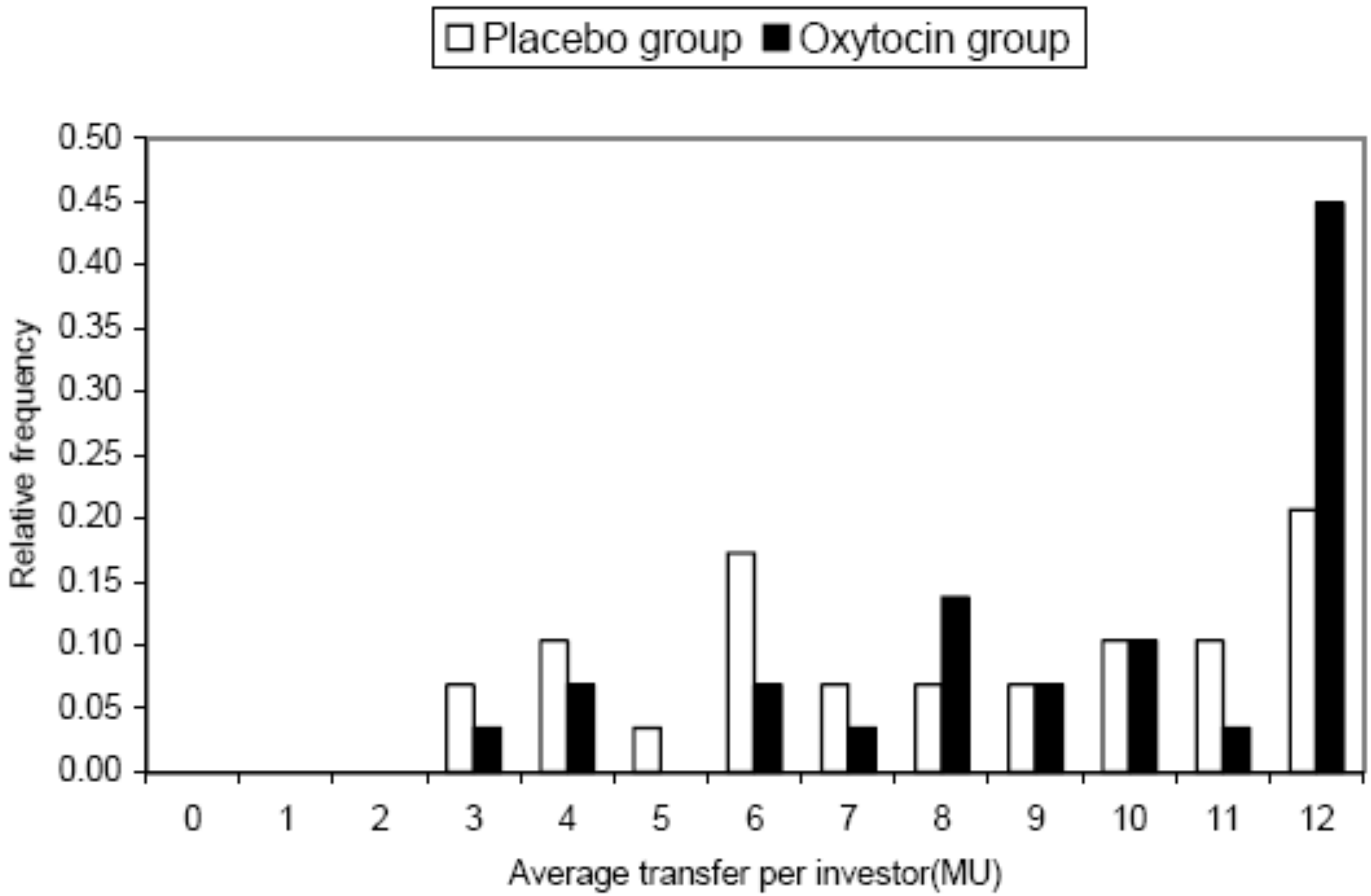
Activation in the caudate nucleus when subjects feel a strong desire to punish others for unfair behavior (compared to control when no such unfair behavior has taken place):



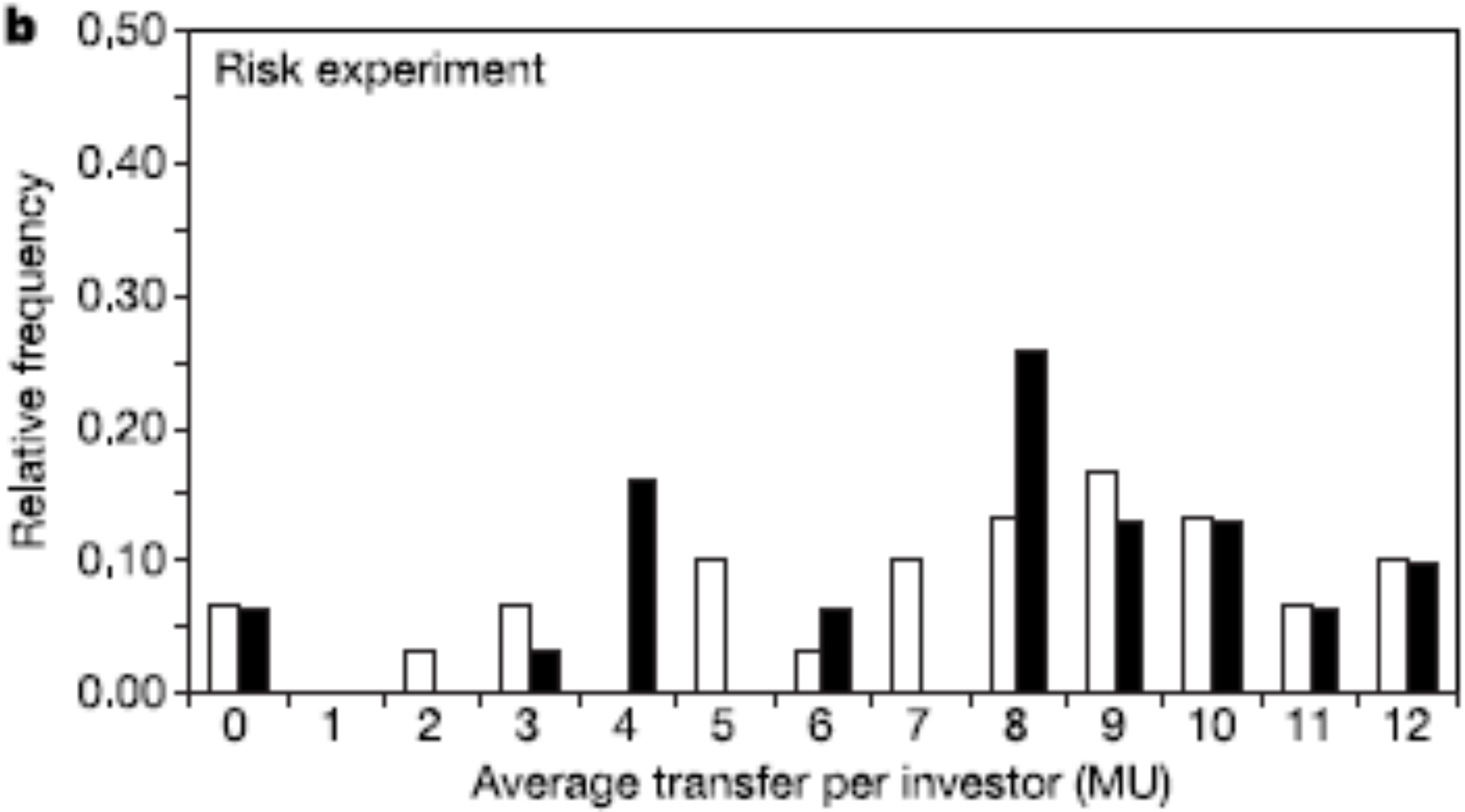
Activation in the prefrontal cortex when subjects know that punishing others will be personally costly to them (compared to control when desire to punish is present but punishment is not costly):



Oxytocin increases trust in humans
(Source: Kosfeld et.al., *Nature*, June 2 2006)



…and it's not about greater willingness to take risks: compare the same game played against a machine…



Conclusions about reason and the emotions

- Evidence from behavioral economics suggests that
  - Individuals care about their self interest and are strategic at pursuing it (and good at anticipating the behavior of others)
  - They also care about the welfare of others (are altruistic)
  - They are also motivated about strong reciprocity, responding to kindness with kindness and to betrayal with revenge
- Evidence from neuroscience suggests that
  - The brain implements cognitive short cuts – such as anatomically separating exploration and exploitation decisions
  - Social preferences (altruism,reciprocity) are anatomically encoded
- How can this be consistent with natural selection?

Other behaviors specific to humans

- As well as depending on social learning, human beings engage in *over-imitation* (studied in particular by Michael Tomasello).
- Children copy as many dimensions of the demonstrators' behavior as they can remember: chimps copy only those whose purpose they can understand. (https://www.youtube.com/watch?v=20Smx_nD9cw)
- In a jungle environment, chimps compete very effectively against children in learning tasks relevant for survival.
- But children can easily beat them when faced with challenges in the laboratory.
- Why would humans have evolved to do this?

Two explanations for the evolution of over-imitation:

- Perhaps the cognitive challenges faced by humans were more difficult, and the solutions less transparent, than those faced by chimps.
- Examples: food extraction from roots, nuts; hunting of large game in time-sensitive environments; warfare (especially ambush).
- In each case, success depends on coordination with others.
- Over-imitation may have other benefits than efficient problem-solving.
- Examples: signaling our commitment to others, our willingness to pay them attention. Evidence that synchrony yields physiological benefits (inc. endorphin release).
- Probably both explanations are partly true. An explanation for ritual?

Toulouse School of Economics

Evolution of Economic Behavior

TSE M1 – Semester 1

September 2023

Paul Seabright

Week 4: The Cognitive and Emotional Foundations of Cooperation.

Economics for the Common Good