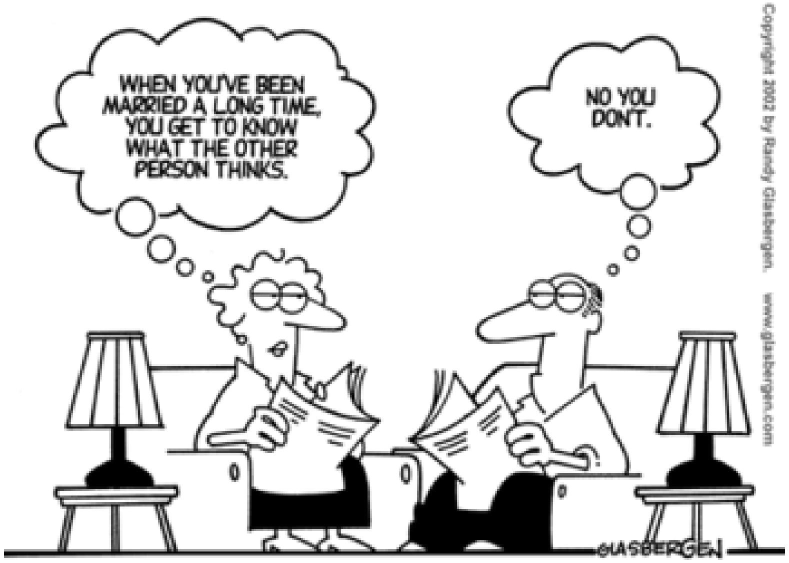
Behavioural, ethological and pathological aspects of ToM: lessons from Bayesian Decision Theory

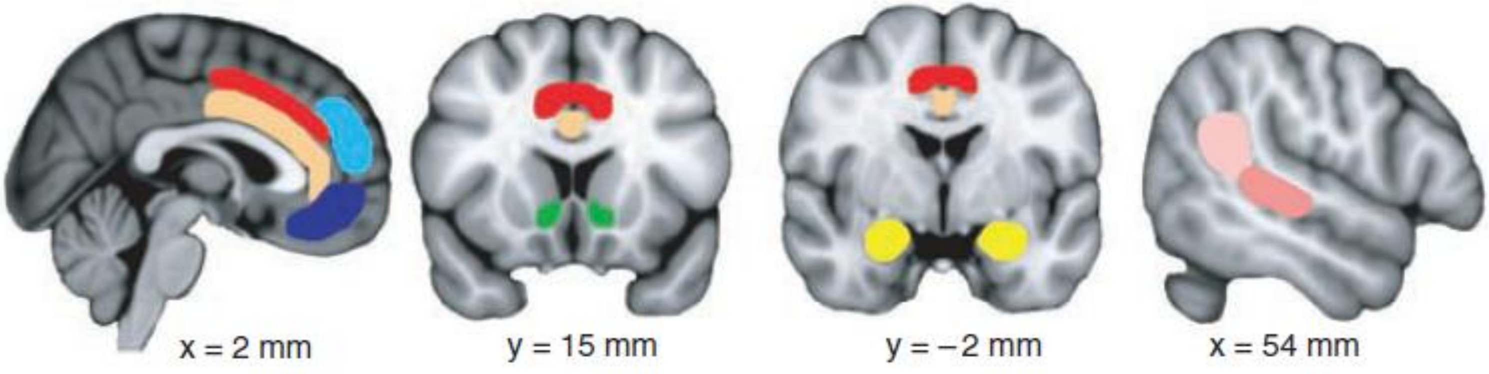


Jean Daunizeau

Brain and Spine Institute (ICM) – Paris, France

ETH – Zurich, Switzerland

The neural bases of social behaviour



- ACCs
- VMPFC
- Amygdala
- VStr

decision making/learning (reward/valuation)

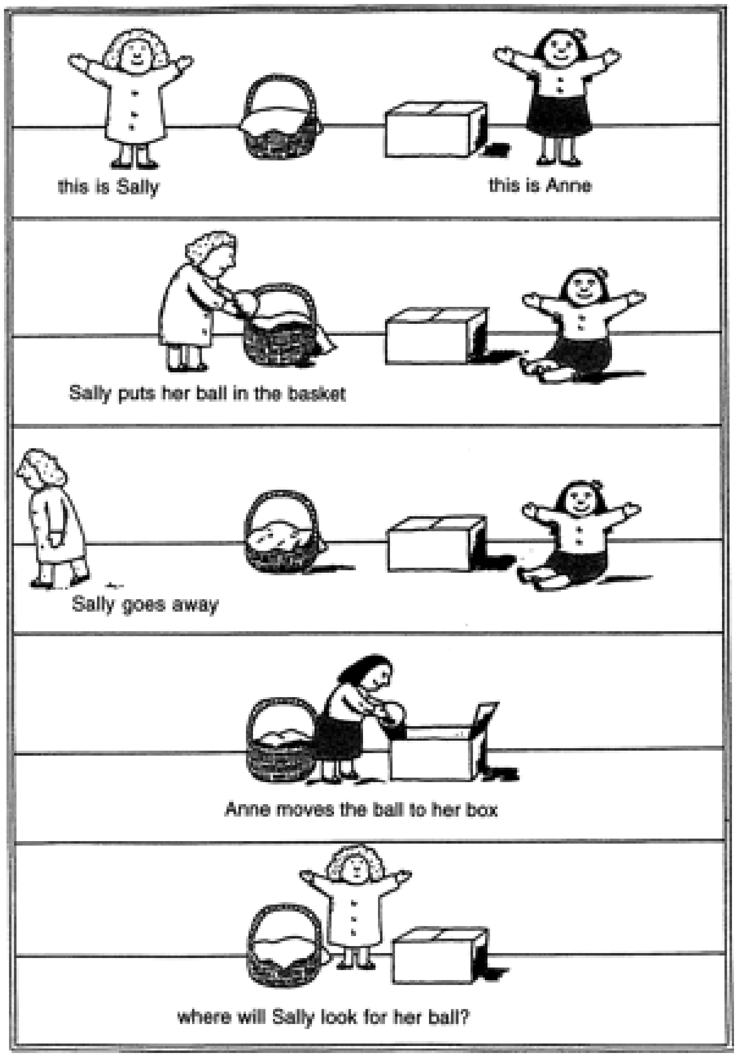
- ACCg
- DMPFC
- TPJ
- STS

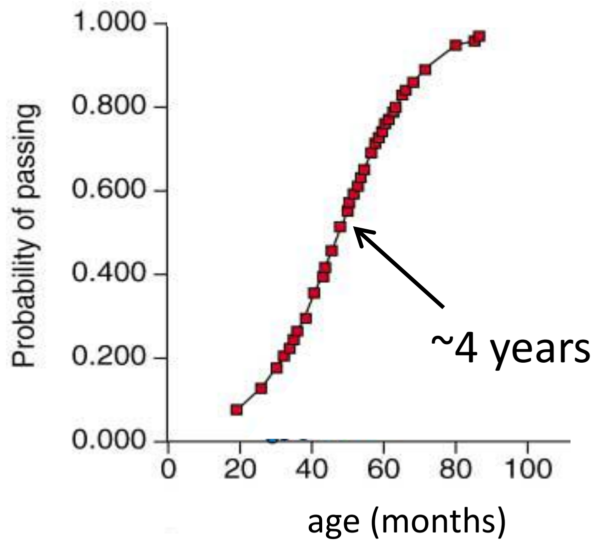
ToM (inferring on the intentions of others)

→ social cognition engage **specific** neural systems

Behrens et al., 2009

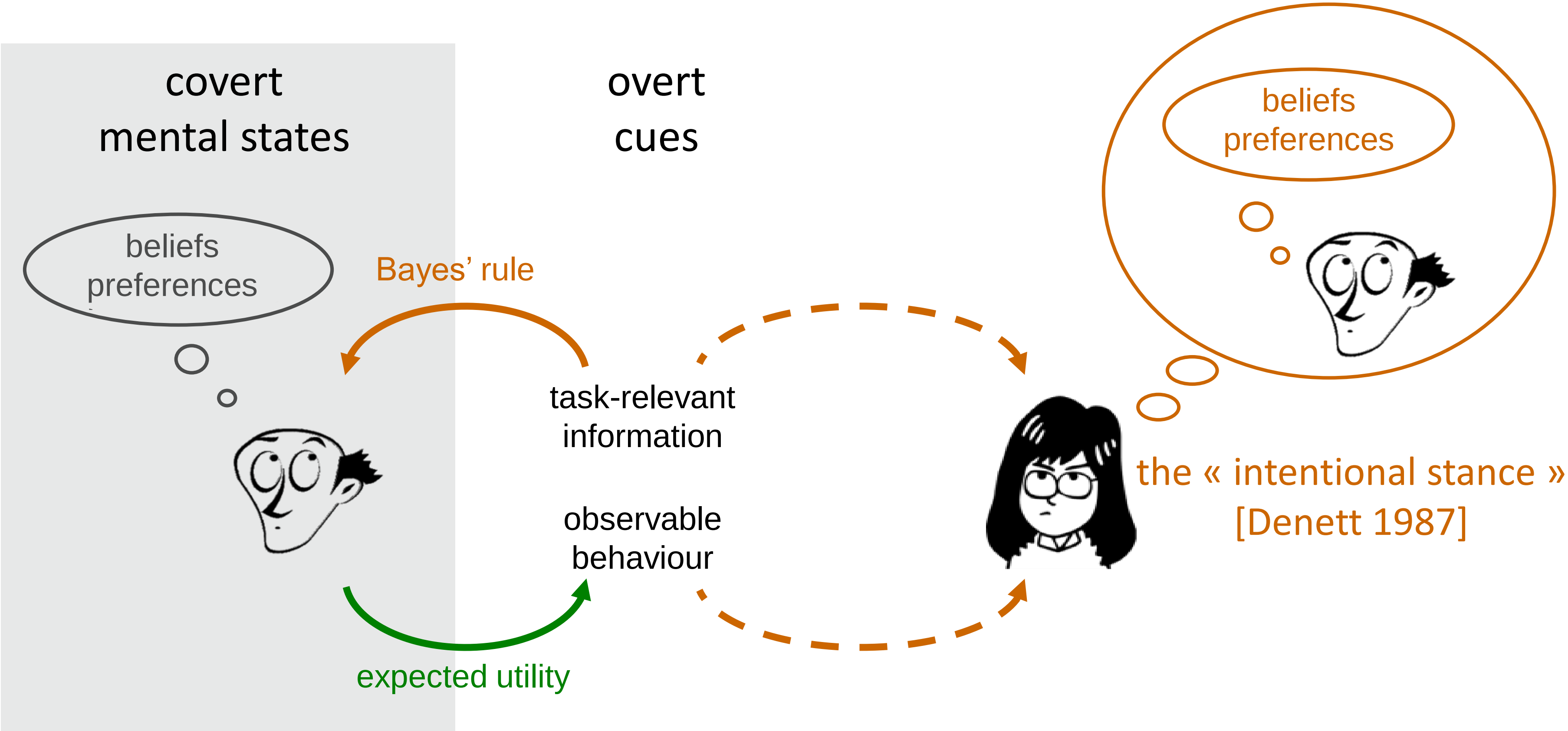
The "false belief" test





Wimmer & Perner, 1983

What computational problem does ToM solve?



ToM = *meta-Bayesian* (Bayesian inference on a Bayesian agent's mental states)?

Overview of the talk

- ✓ Does ToM make a difference when we learn?
- ✓ Limited ToM sophistication: did evolution fool us?
- ✓ Playing *hide-and-seek* with non-human primates
- ✓ What about people with autism spectrum disorder?

Overview of the talk

- ✓ Does ToM make a difference when we learn?
- ✓ Limited ToM sophistication: did evolution fool us?
- ✓ Playing *hide-and-seek* with non-human primates
- ✓ What about people with autism spectrum disorder?

0-ToM

0-ToM does not apply the intentional stance

→ 0-ToM is a Bayesian agent with:

- beliefs (about non-intentional contingencies)
- preferences

« I believe that you will hide behind the tree »

1-ToM

1-ToM learns how the other learns

→ 1-ToM is a meta-Bayesian agent with:

- beliefs (about other's beliefs and preferences)

- preferences

« I believe that you believe that I will hide behind the tree »

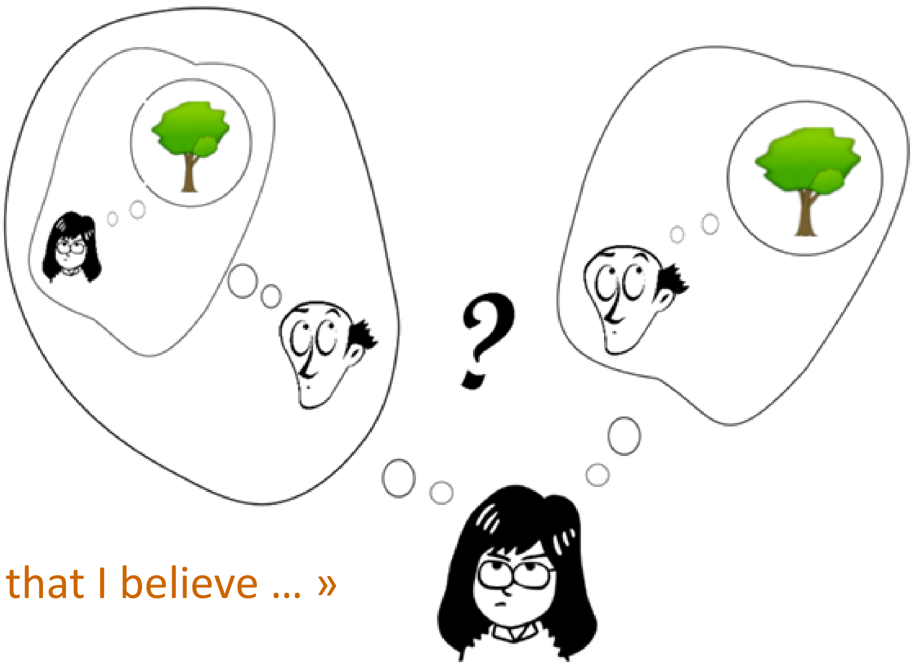
2-ToM

2-ToM learns how the other learns and her ToM sophistication level

→ 2-ToM is a meta-Bayesian agent with:

- beliefs (about other's beliefs – about one's beliefs - and preferences)

- preferences



« I believe that you believe that I believe … »

k-ToM: recursive meta-Bayesian modelling

- k-ToM learns how the other learns and her ToM sophistication level:

$$\lambda_{\tau}^{(k)} = f\left(\lambda_{\tau-1}^{(k)}, a_{\tau}, \theta_{1}^{(k)}\right)$$

- k-ToM acts according to her beliefs and preferences:

$$p\left(a_{1,\tau+1} \middle| \theta^{(k)}\right) \propto \exp - \rho\left(\lambda_{\tau+1}^{(k)}, a_{1,\tau+1}\right) \Big/ \theta_{2}^{(k)}$$

- This induces a likelihood for a k+1-ToM observer:

$$p\left(a_{1,\to\tau} \middle| \theta^{(1,\ldots,k)}, \kappa, m_{k+1}\right) = \prod_{k'=0}^{k} \prod_{\tau'=1}^{\tau} p\left(a_{1,\tau'} \middle| \theta^{(k)}\right)^{\zeta_{k'}(\kappa)}$$

- Deriving the ensuing Free-Energy yields the k+1-ToM learning rule:

$$\lambda_{\tau+1}^{(k+1)} = f\left(\lambda_{\tau}^{(k+1)}, a_{\tau}, \theta_{1}^{(k+1)}\right)$$

$$f : \lambda_{\tau}^{(k+1)} \to \arg\max_{\lambda_{\tau+1}^{(k+1)}} F_{\tau}^{(k+1)}$$

$$F_{\tau}^{(k+1)} = \left\langle \ln p\left(a_{1,\to\tau} \middle| \theta^{(1,\ldots,k)}, \kappa, m_{k+1}\right)\right\rangle + \left\langle \ln p\left(\theta^{(1,\ldots,k)}, \kappa \middle| m_{k+1}\right)\right\rangle - \left\langle \ln q_{\tau}\left(\theta^{(1,\ldots,k)}, \kappa\right)\right\rangle$$

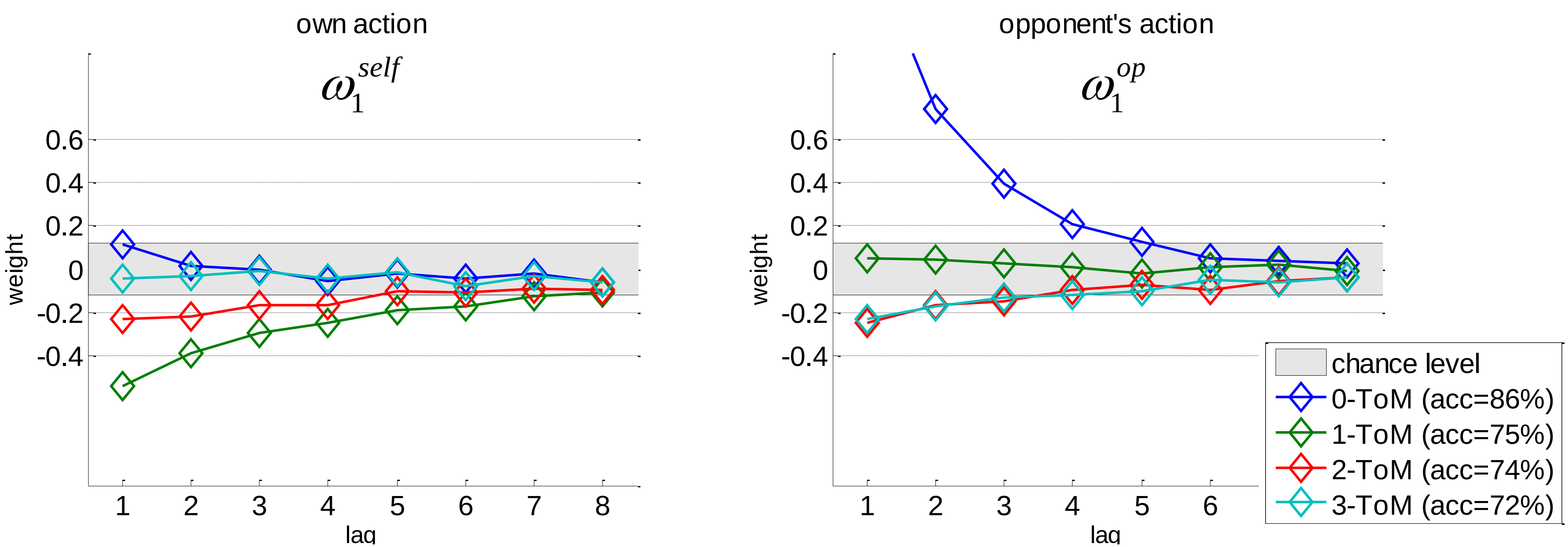
k-ToM agents in competitive games

outcome table
(« hide and seek »)

	hider: $a_1 = 1$	hider: $a_1 = 0$
seeker: $a_2 = 1$	-1, 1	1, -1
seeker: $a_2 = 0$	1, -1	-1, 1

Volterra 1st-order kernels:

$$p\left(a_t = 1 \middle| \omega\right) = s\left(\omega_0 + \sum_k \sum_\tau \omega_\tau^{(k)} u_{t-\tau}^{(k)} + \ldots\right)$$



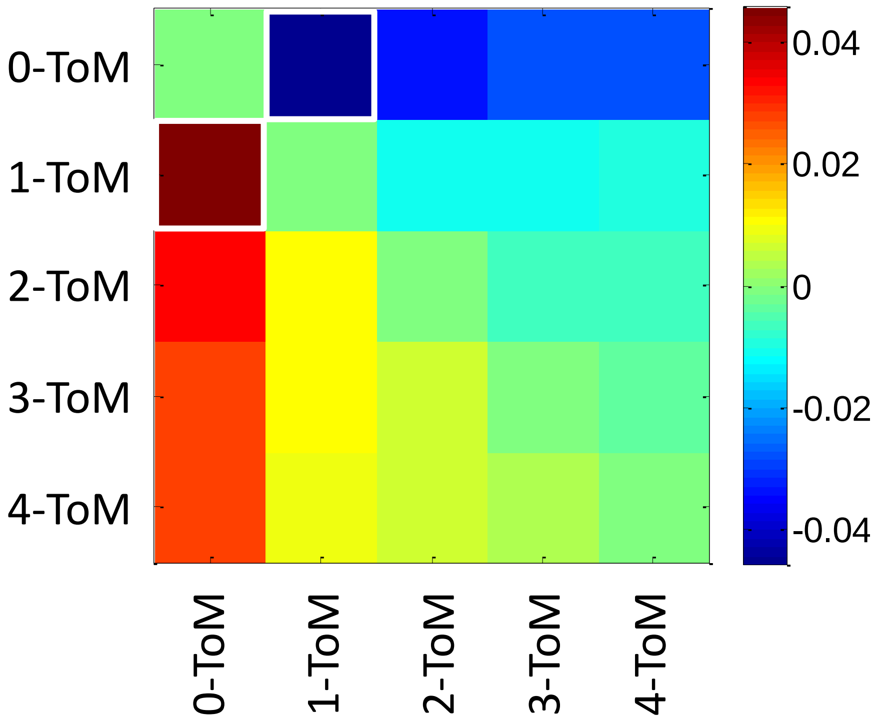
k-ToM agents in competitive games

outcome table
(« hide and seek »)

	hider: a_1 = 1	hider: a_1 = 0
seeker: a_2 = 1	-1, 1	1, -1
seeker: a_2 = 0	1, -1	-1, 1

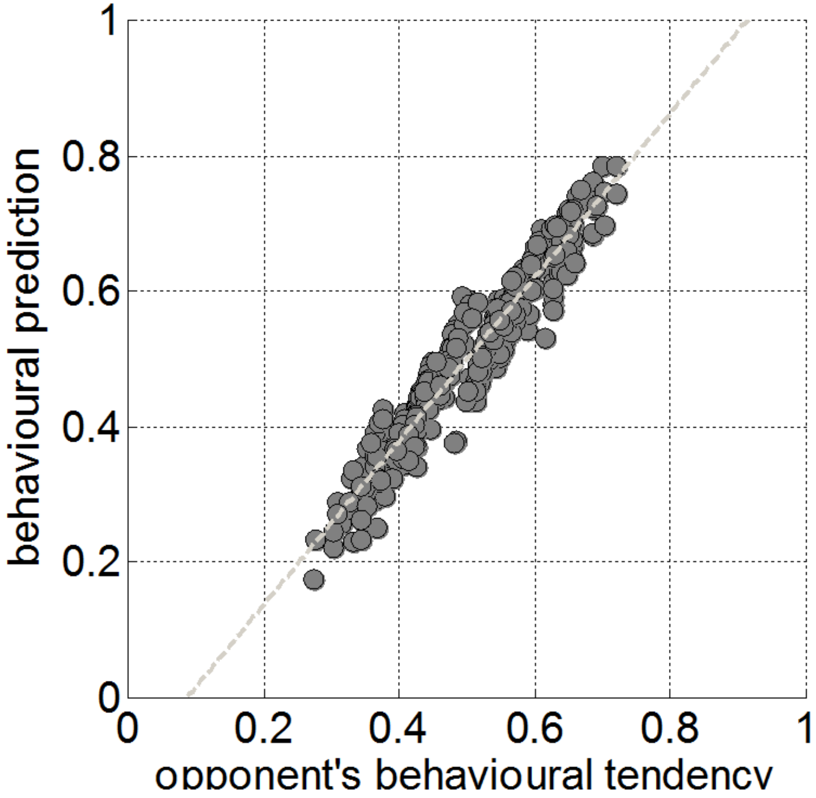
simulated
behavioural performance
(#wins/trial)

$$\tau = 512$$

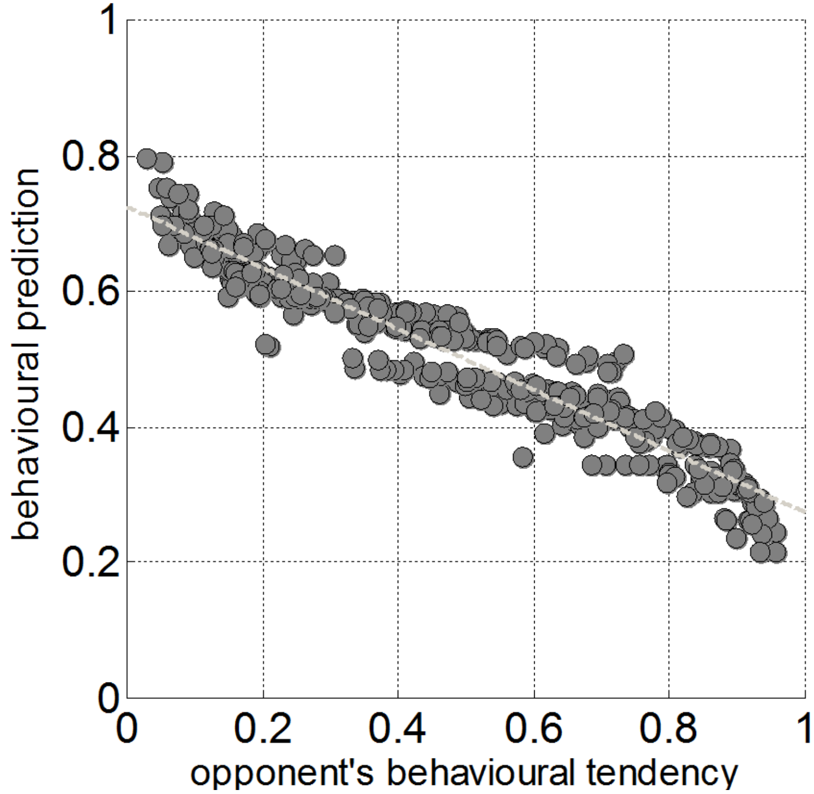


Everybody is somebody's fool

1-ToM predicts 0-ToM



0-ToM predicts 1-ToM



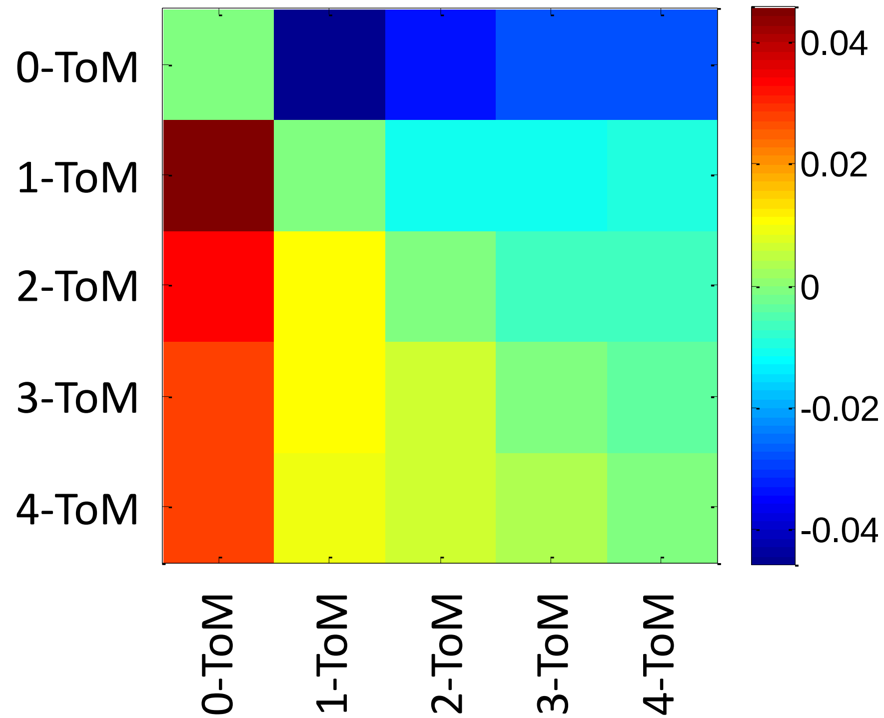
k-ToM agents in competitive games

outcome table
(« hide and seek »)

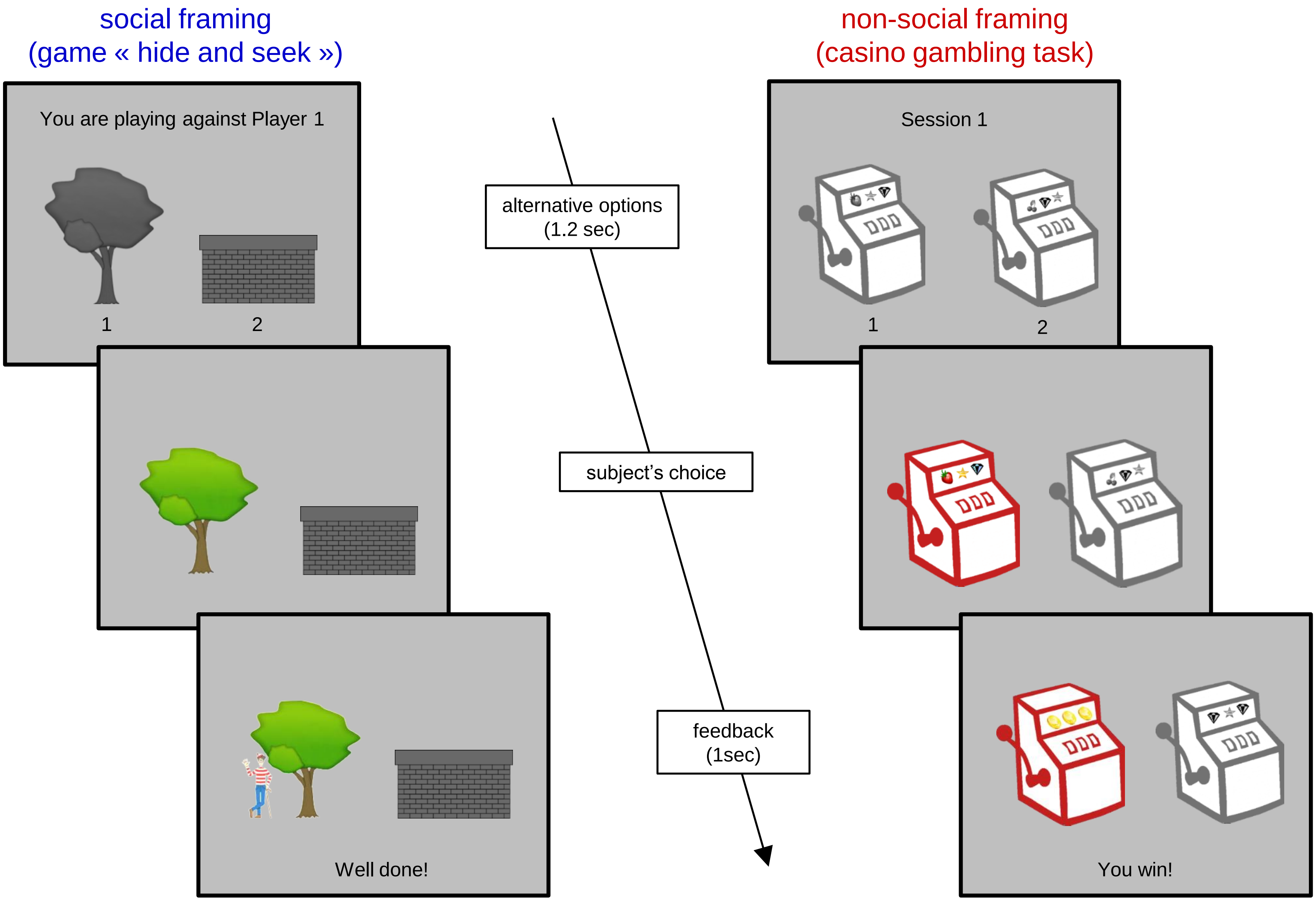
	hider: $a_1 = 1$	hider: $a_1 = 0$
seeker: $a_2 = 1$	-1, 1	1, -1
seeker: $a_2 = 0$	1, -1	-1, 1

simulated
behavioural performance
(#wins/trial)

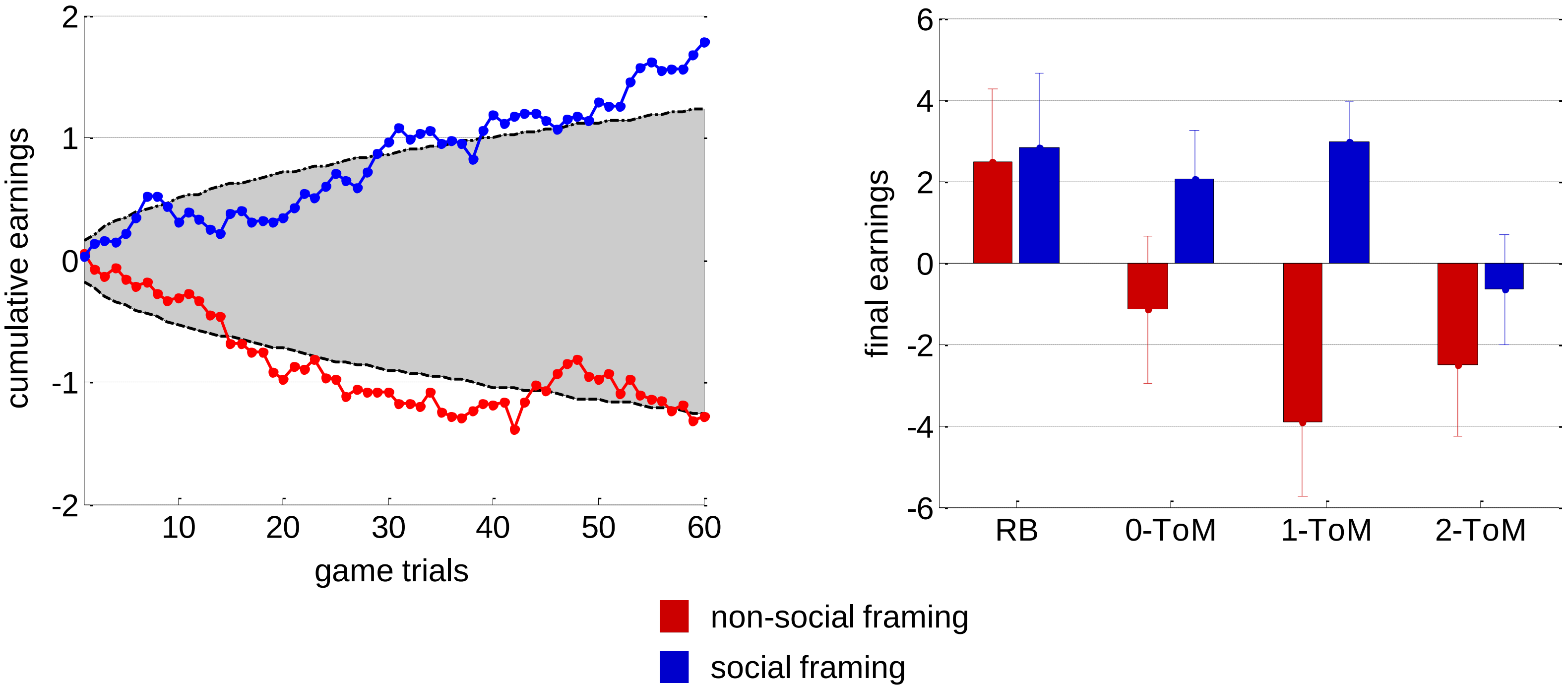
$$\tau = 512$$



Behavioural task design



Behavioural performances (N=26)



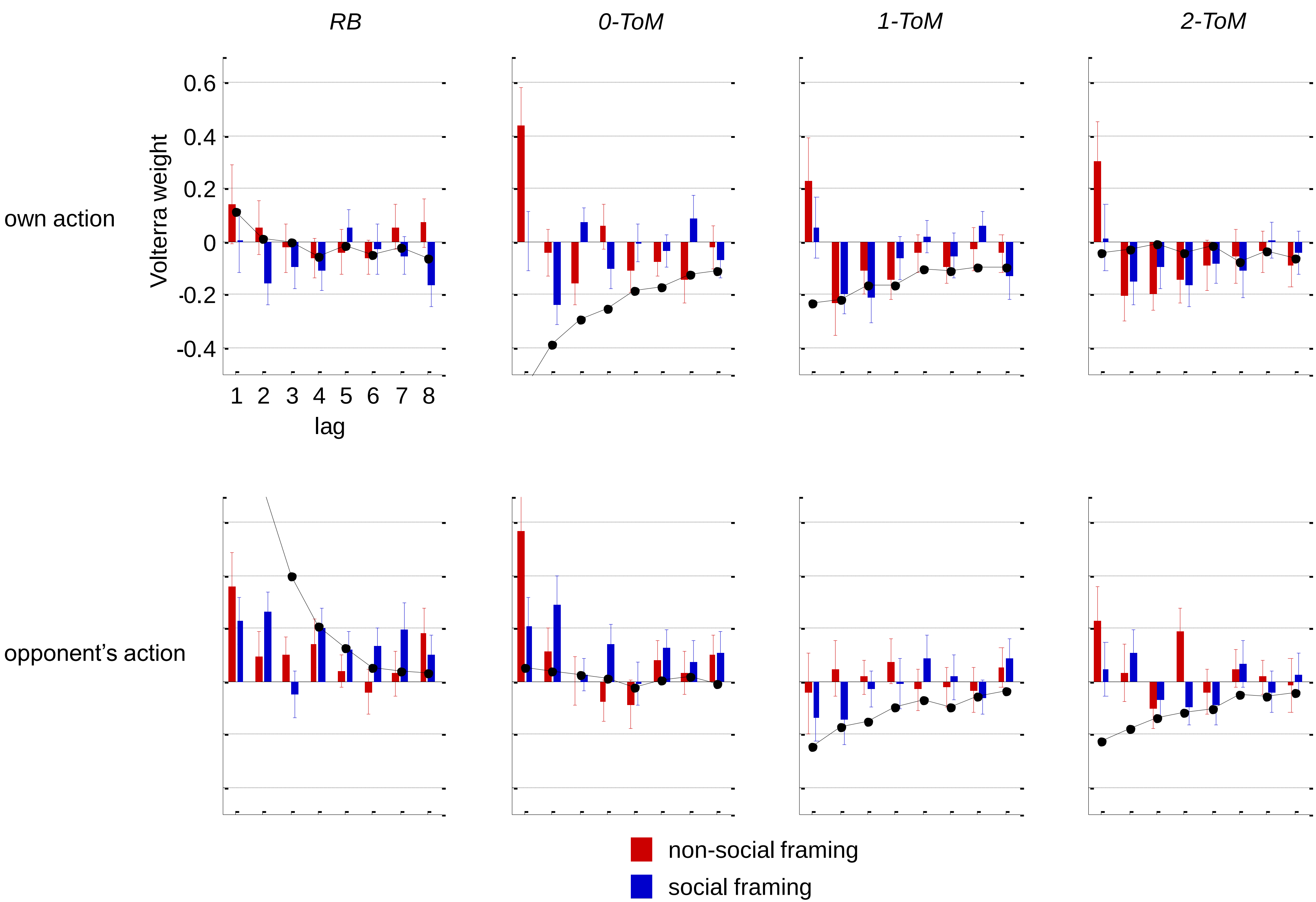
ANOVA :

- **Framing** (p=0.007), **opponent** (p=0.009), 0 framingXop (but RB VS 1-ToM)
- 0 age, 0 sex

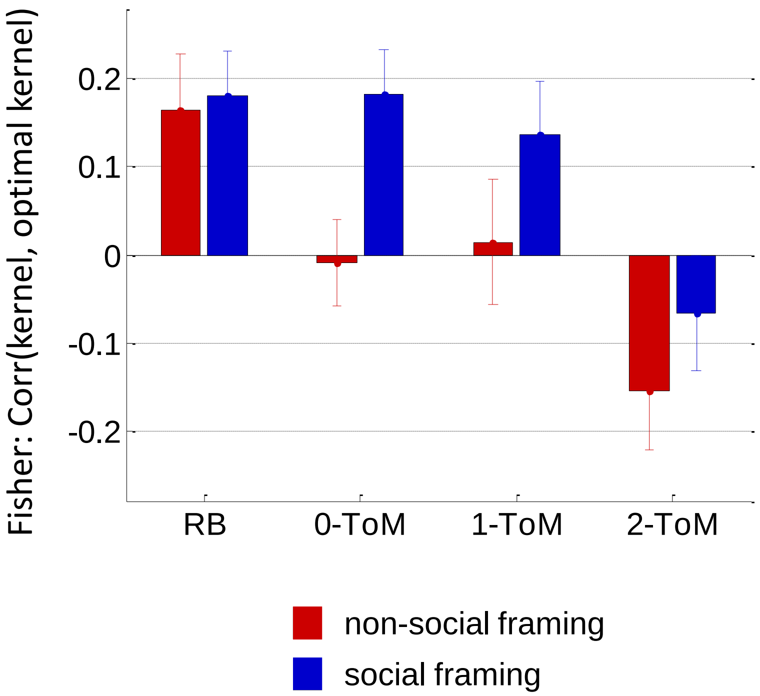
BETWEEN-SUBJECT VARIABILITY:

- 0 empathy, 0 executive functions (WCST, Go-NoGo, 3-back)
- **RB: Corr NS & S** (p=0.01), 0 otherwise

Volterra decompositions



Similarity to best k-ToM response



ANOVA:

- **Framing** (p=0.02), **op** (p=0.0001), 0 framingXop
- 0 age, 0 sex

SOBEL:

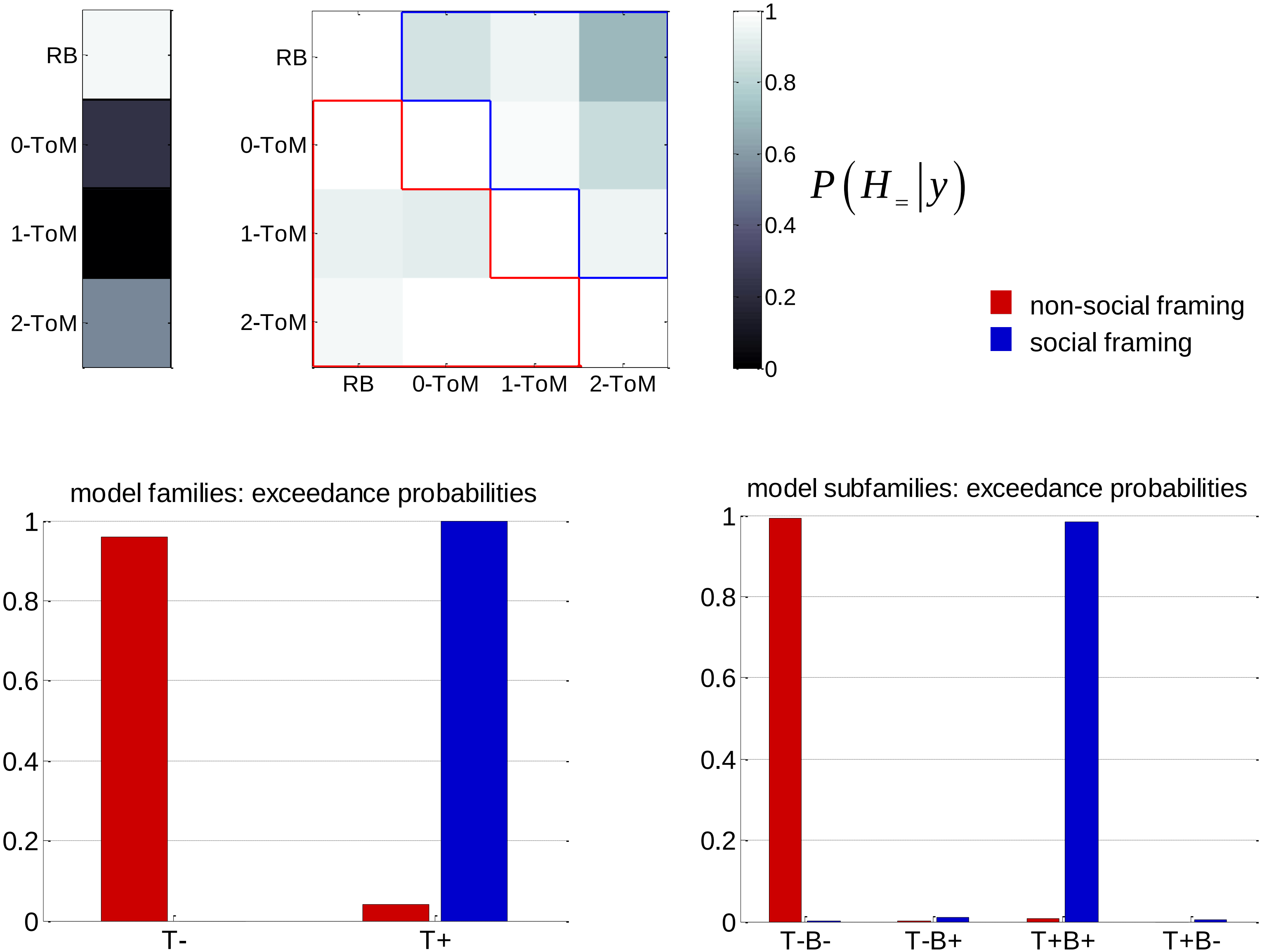
- **mediation of framing** (p=0.010), **mediation of op** (p=0.013)

Bayesian model comparison

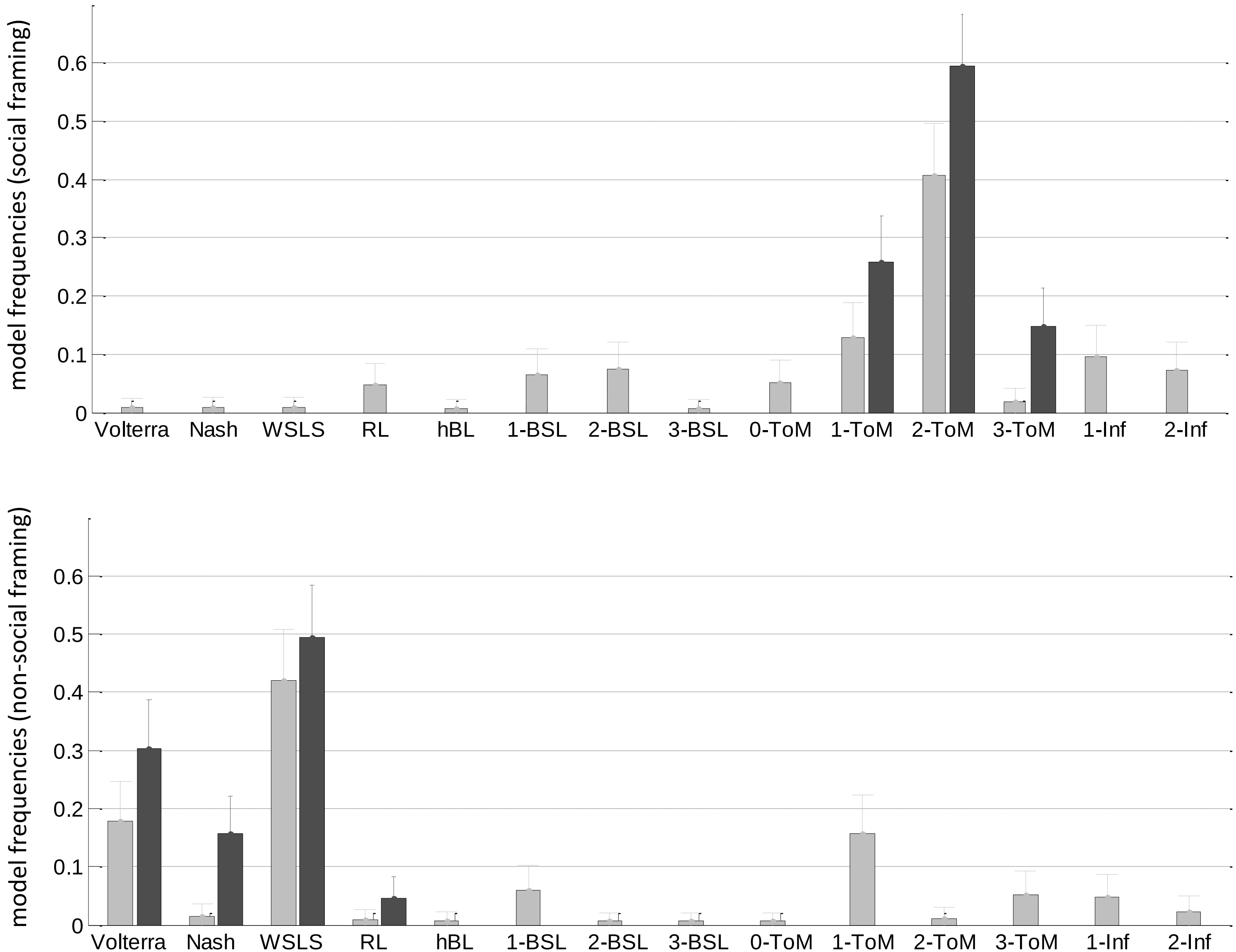
Model's name	Bayesian	mentalizing	number of free parameters
k-ToM $(1 \le k \le 3)$	yes (B+)	yes (T+)	3
0-ToM	yes (B+)	no (T-)	3
HGF	yes (B+)	no (T-)	5
n-BSL $(1 \le n \le 3)$	yes (B+)	no (T-)	3
k-Inf $(1 \le k \le 2)$	no (B-)	yes (T+)	3 (*1-Inf*), 4 (*2-Inf*)
RL	no (B-)	no (T-)	3
WSLS	no (B-)	no (T-)	2
Nash	no (B-)	no (T-)	1

- 14 models, 26 participants, 2 tasks framings, 4 opponents (= 2912 model evidences)
- 2X2 model families (2 partitions: B+/B-, T+/T-)

Bayesian model comparison



Variability of human ToM sophistication



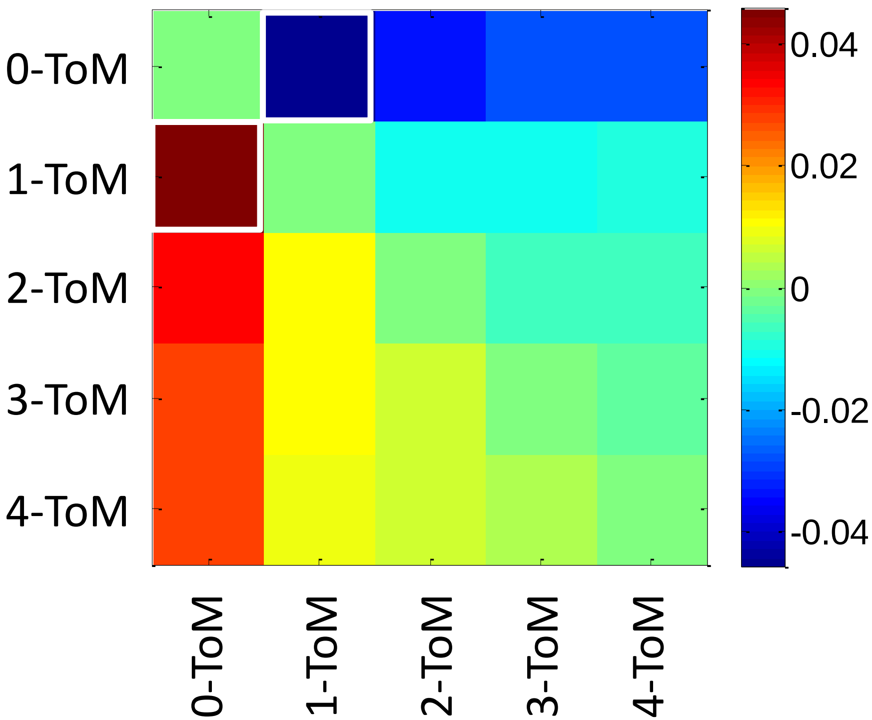
Overview of the talk

- ✓ Does ToM make a difference when we learn?
- ✓ Limited ToM sophistication: did evolution fool us?
- ✓ Playing *hide-and-seek* with non-human primates
- ✓ What about people with autism spectrum disorder?

Competitive versus cooperative games

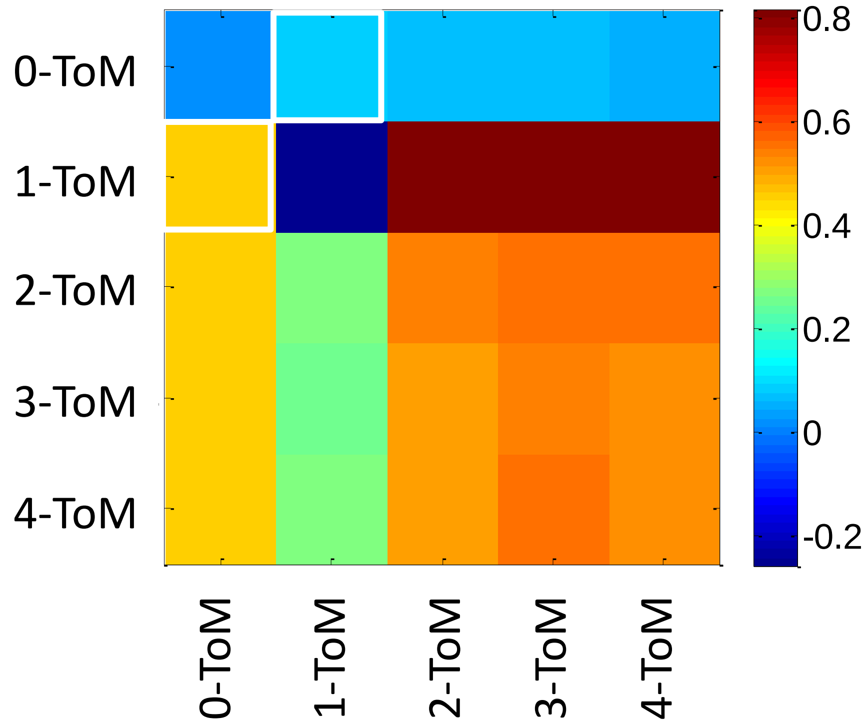
« hide and seek »

	P1: a_1 = 1	P1: a_1 = 0
P2: a_2 = 1	-1, 1	1, -1
P2: a_2 = 0	1, -1	-1, 1



« battle of the sexes »

	P1: a_1 = 1	P1: a_1 = 0
P2: a_2 = 1	2, 0	-1, -1
P2: a_2 = 0	-1, -1	0, 2



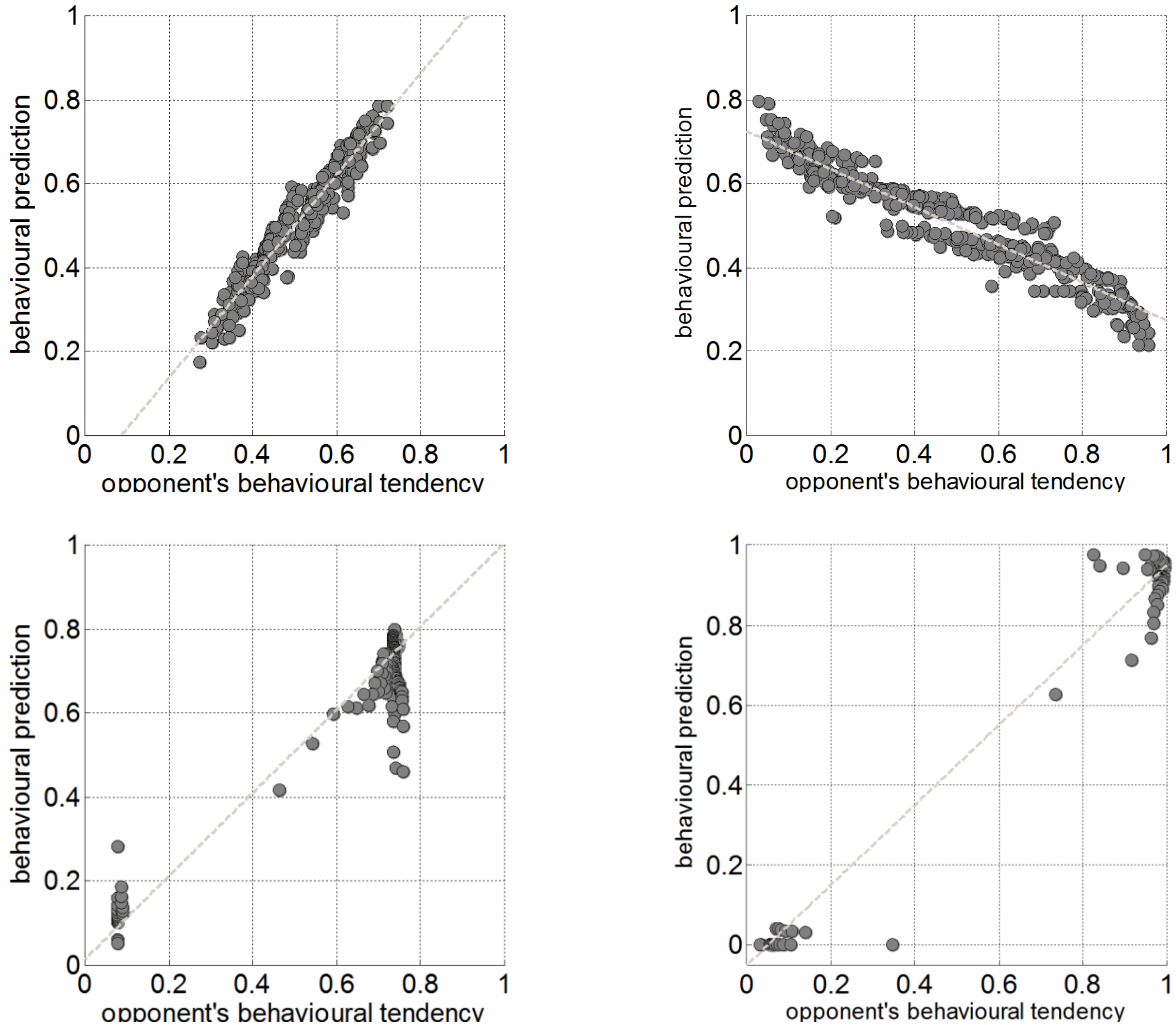
Being right is as good as being smart

1-ToM predicts 0-ToM

0-ToM predicts 1-ToM

« hide and seek »

« battle of the sexes »



Evolutionary game theory

Can we explain the emergence of the natural bound on ToM sophistication?

→ Average adaptive fitness:

- is a function of the behavioural performance, relative to other phenotypes
- depends upon the frequency of other phenotypes within the population

s_k	frequency of phenotype *k* within the population
ω_i	frequency of game *i*
$Q^{(i)}(\tau)$	expected payoff matrix of game *i* at round τ

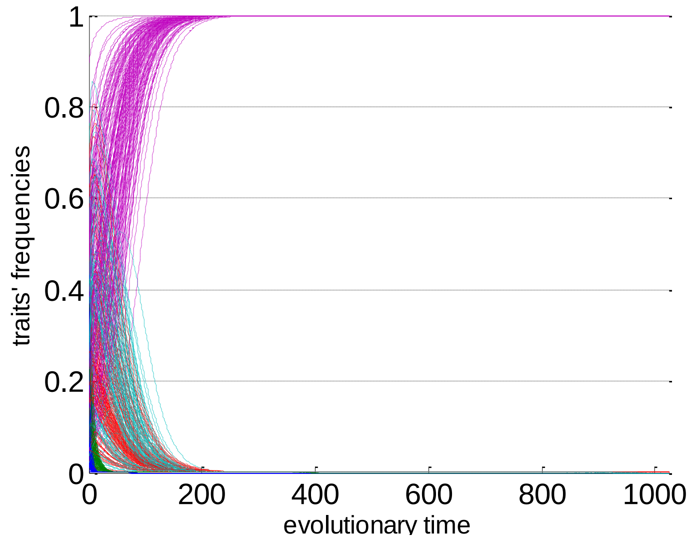
→ Replicator dynamics [Maynard-Smith 1982, Hofbauer 1998]:

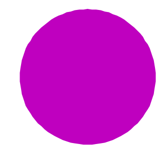
$$\frac{ds}{dt} = Diag(s)\left(\sum_i \omega_i\, Q^{(i)}(\tau)s - \sum_i \omega_i\, s^T Q^{(i)}(\tau)s\right)$$

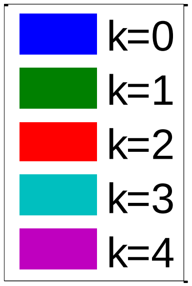
evolutionary stable states: $s_\infty \equiv \lim_{t\to\infty} s(t)$

Replicator dynamics and ESS

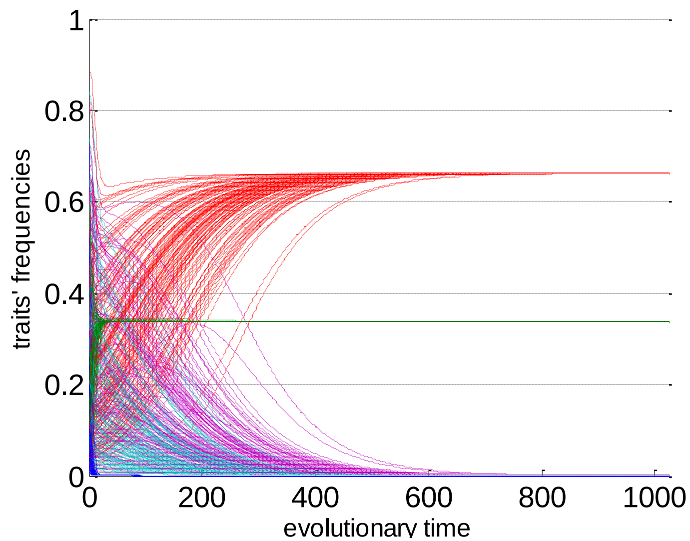
« hide and seek »

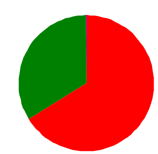




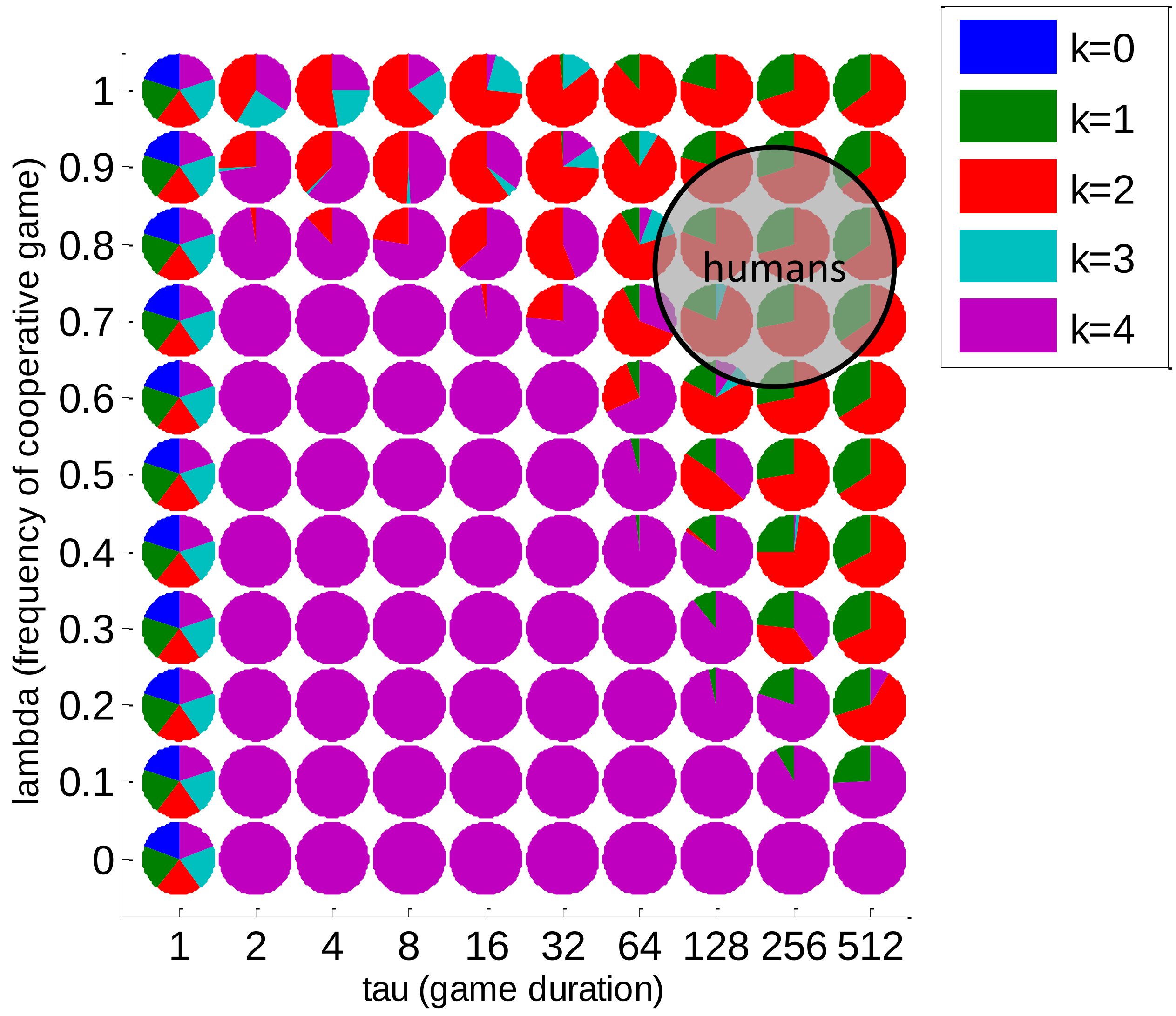


« battle of the sexes »





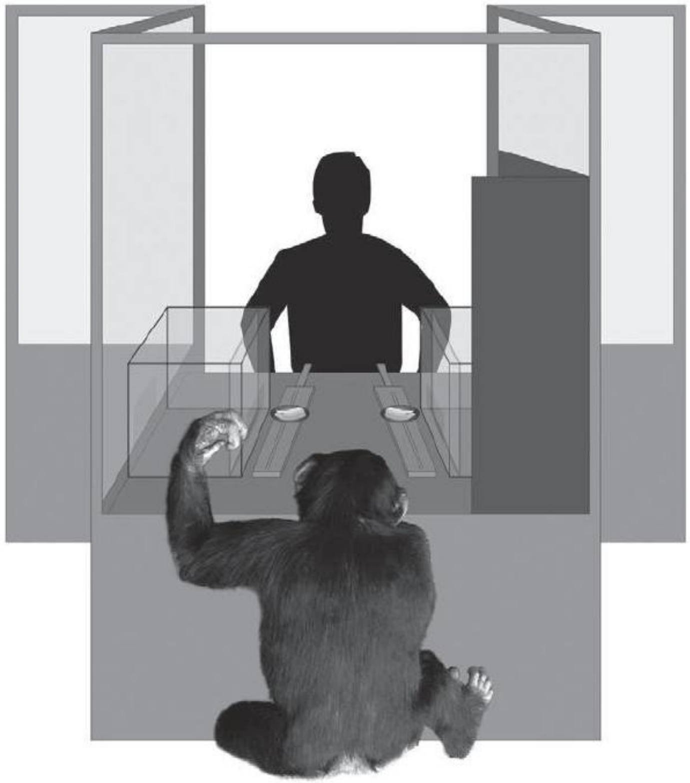
ESS: phase portrait



Overview of the talk

- ✓ Does ToM make a difference when we learn?
- ✓ Limited ToM sophistication: did evolution fool us?
- ✓ Playing *hide-and-seek* with non-human primates
- ✓ What about people with autism spectrum disorder?

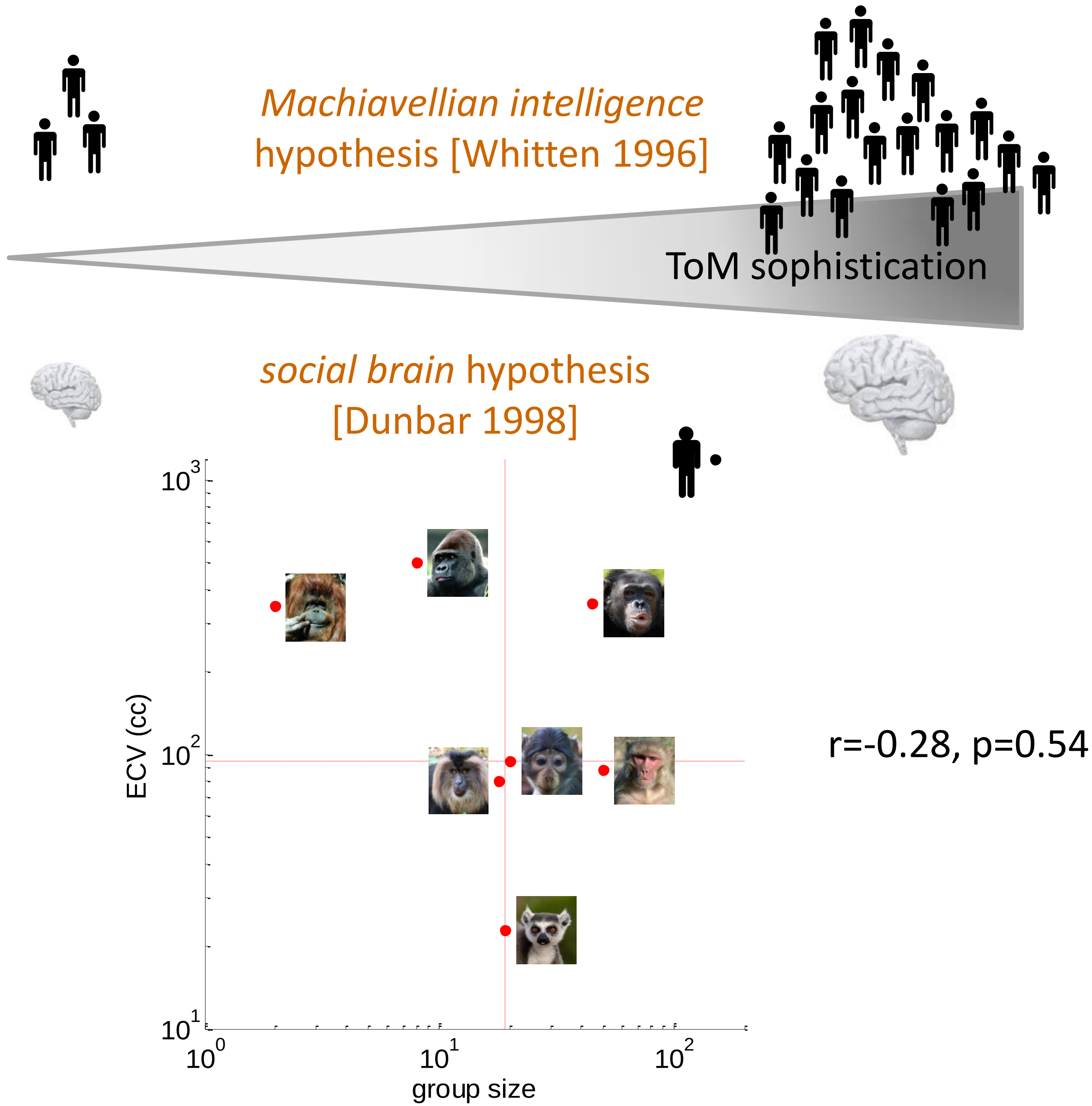
The main confound in primates' ToM assessment



You're competing for the food.
Where should you approach the food from?

[Hare 2006]

Evolutionary factors of ToM sophistication

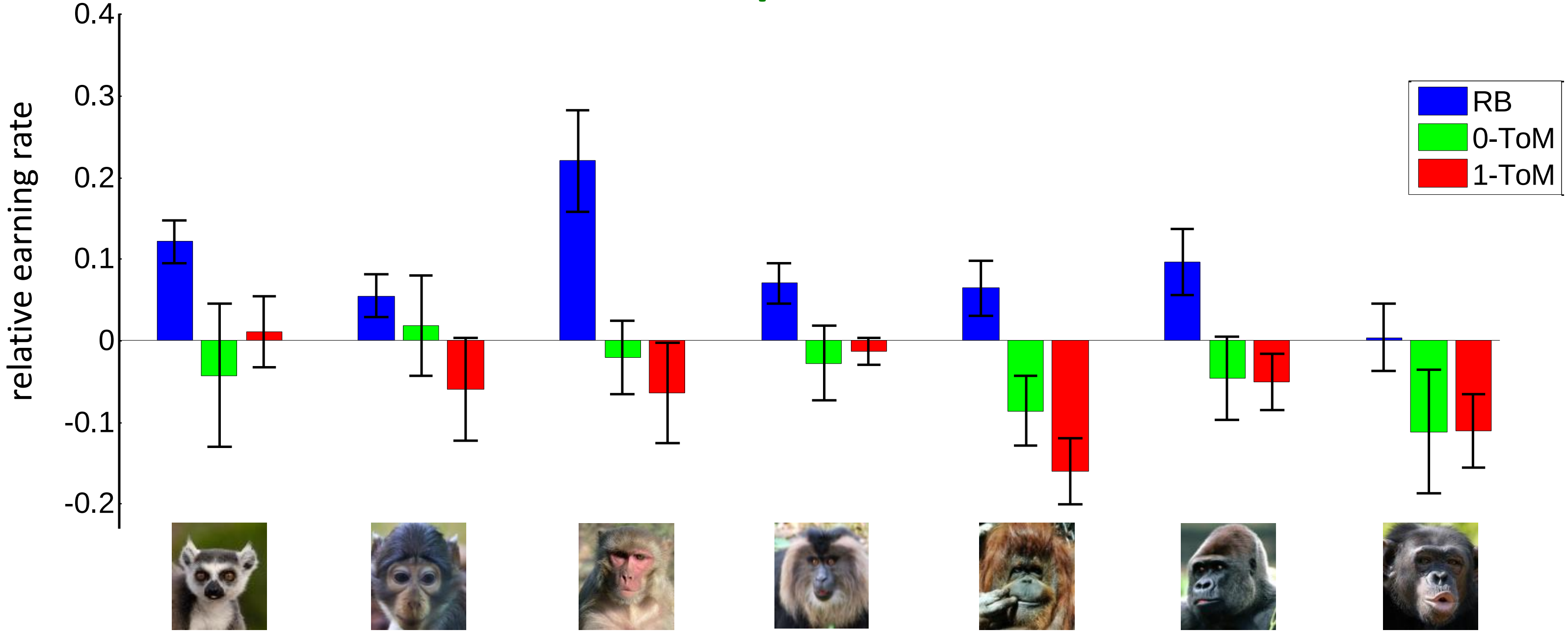


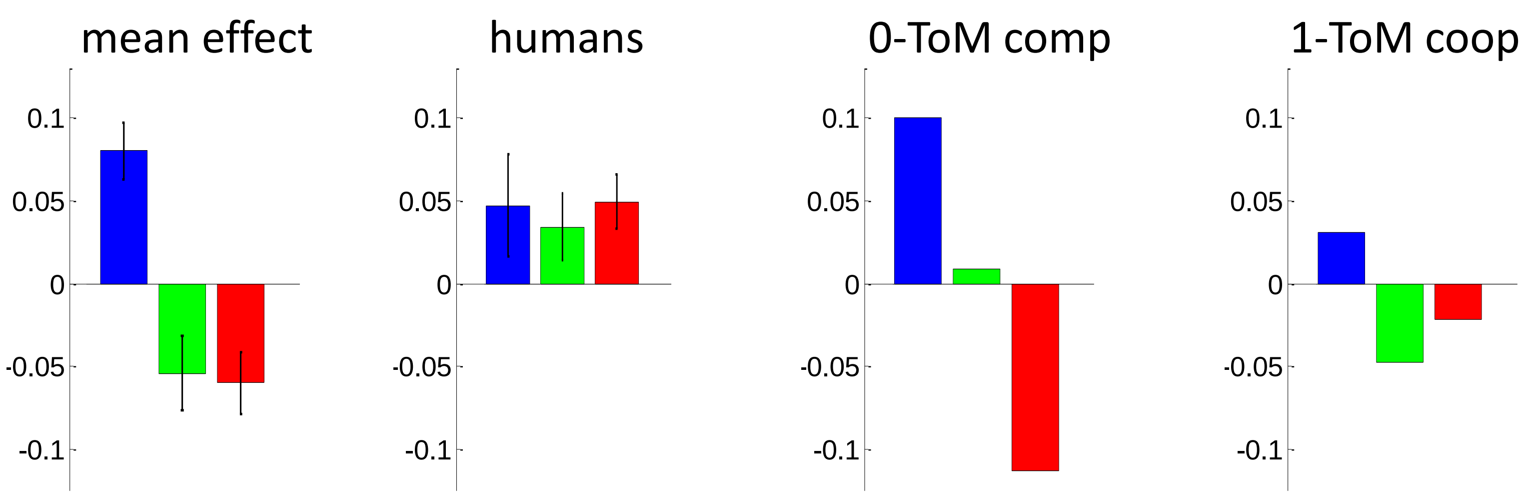
Playing “hide and seek“ with primates

• ***Subjects:***
Macaques (4+5), Orangutans (7), Chimps (6), Gorillas (5), Mangabeys (8), Lemurs (6)

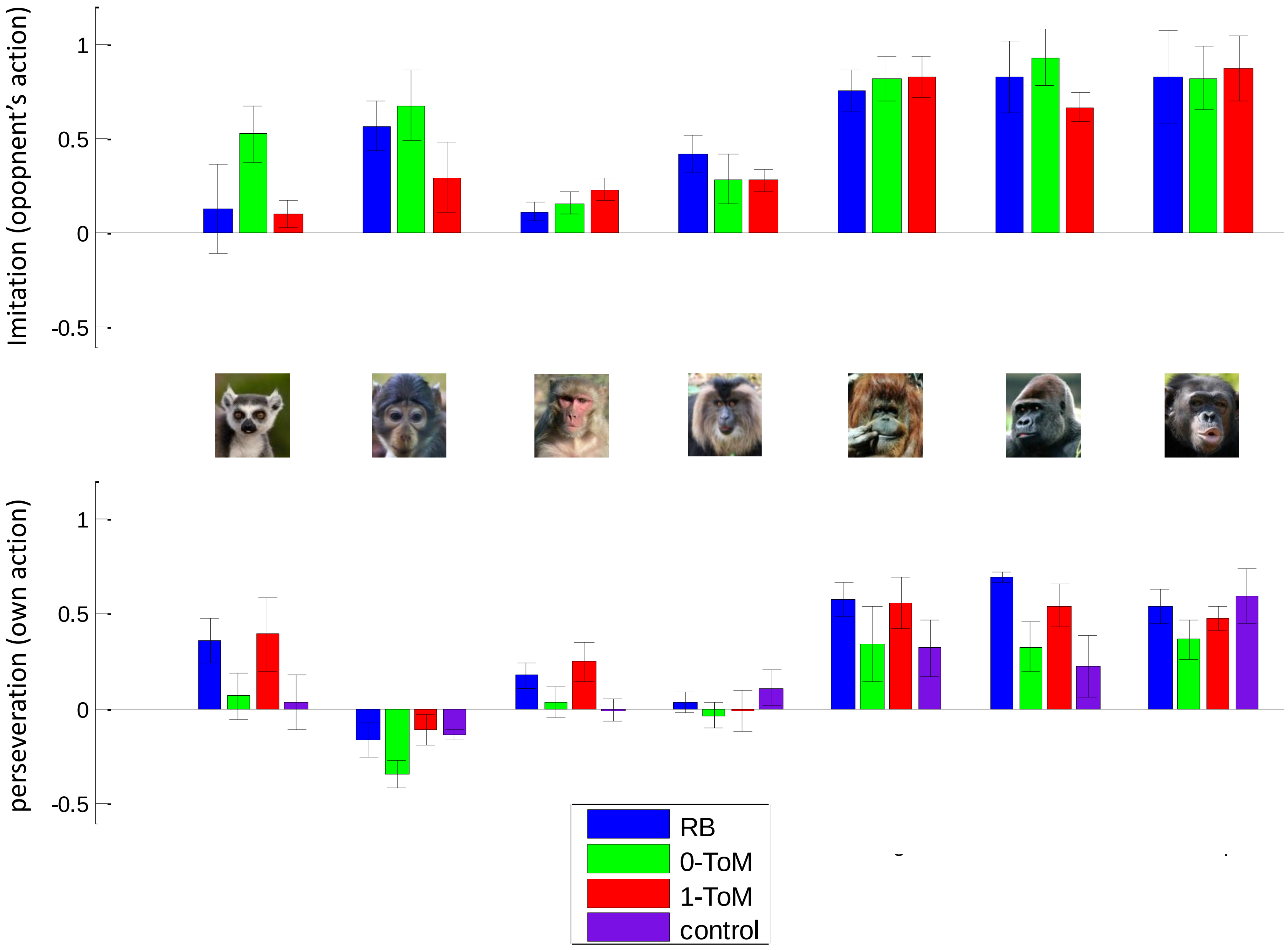
• ***Experimental paradigm:***
- ✓ habituation/training sessions (rule learning)
- ✓ 3 opponent types (RB, 0-ToM, 1-ToM) X 4 sessions
- ✓ control task (behavioural perseveration)

Behavioural performances

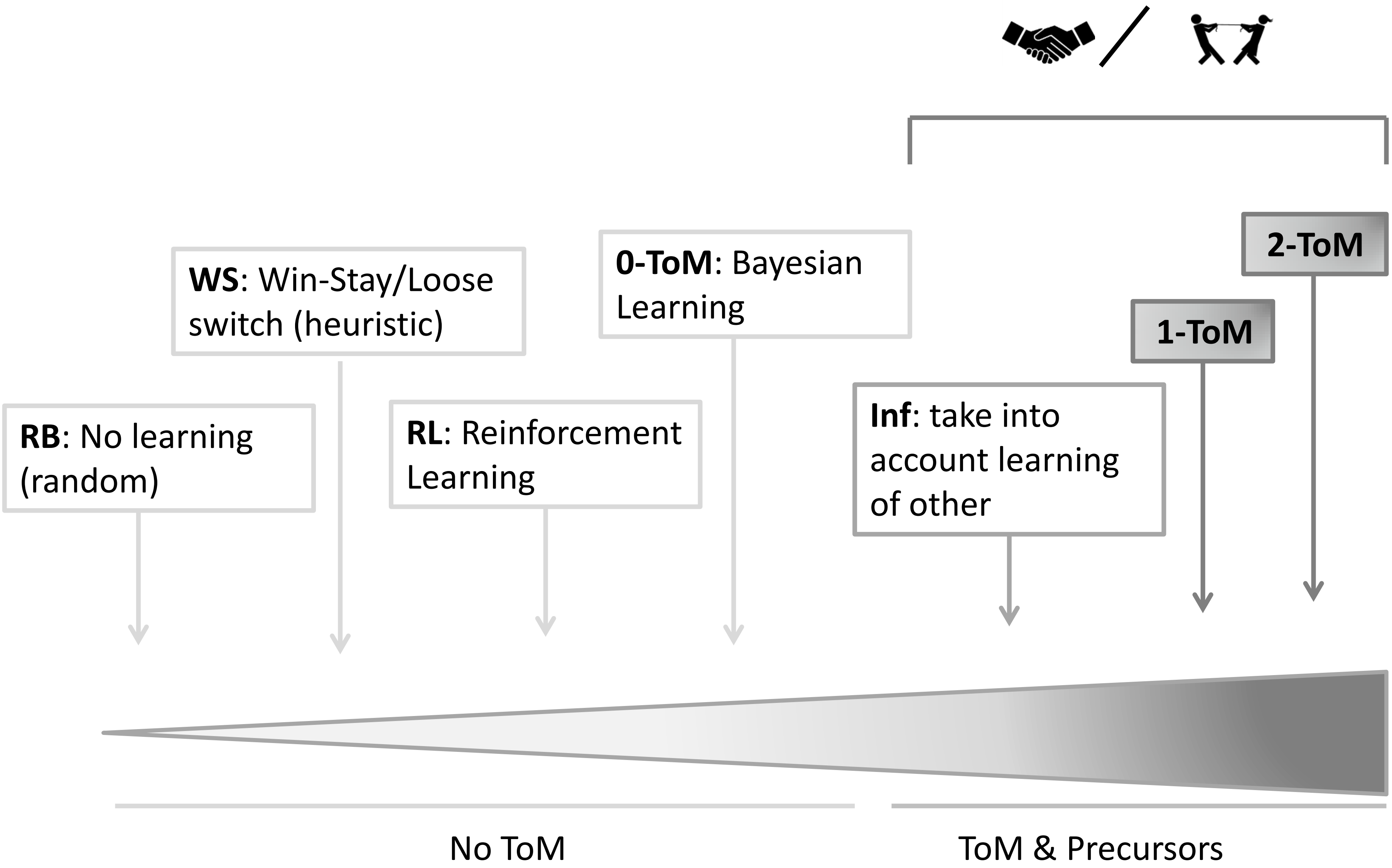




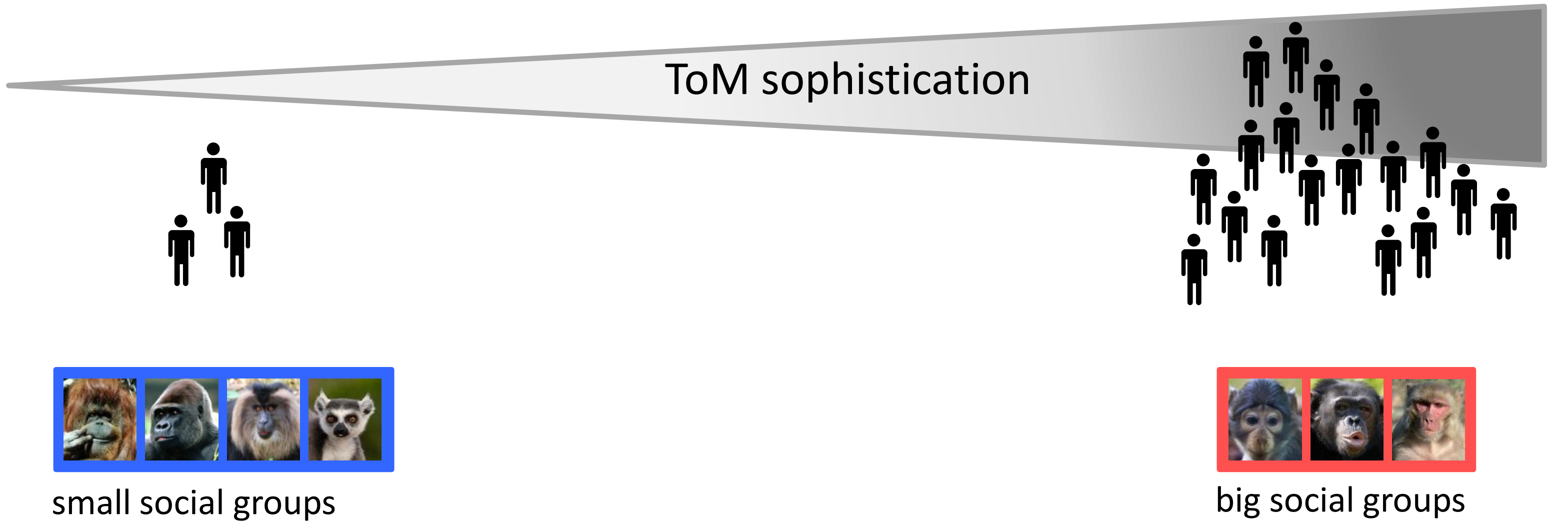
Volterra decompositions

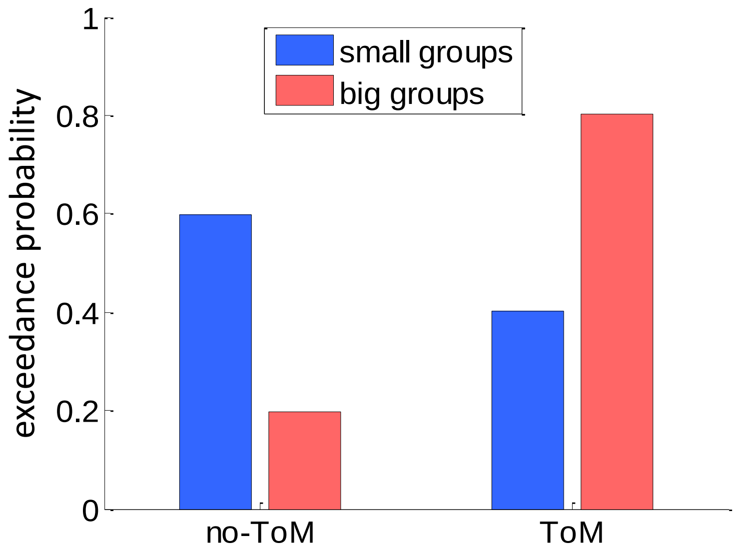


ToM sophistication of learning styles

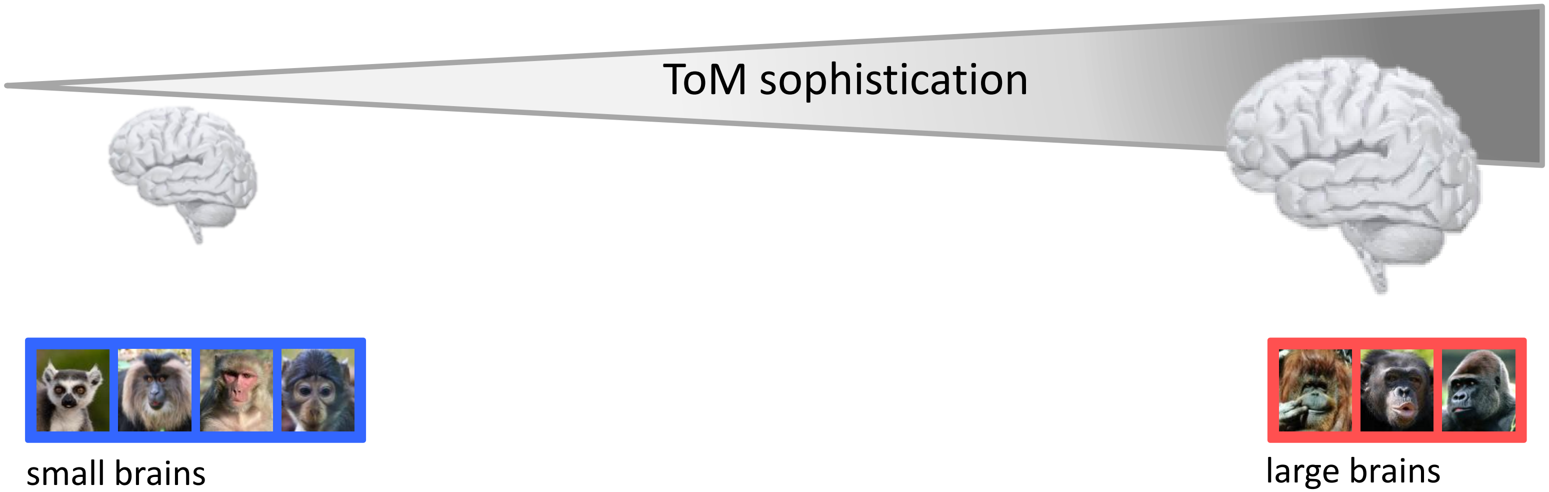


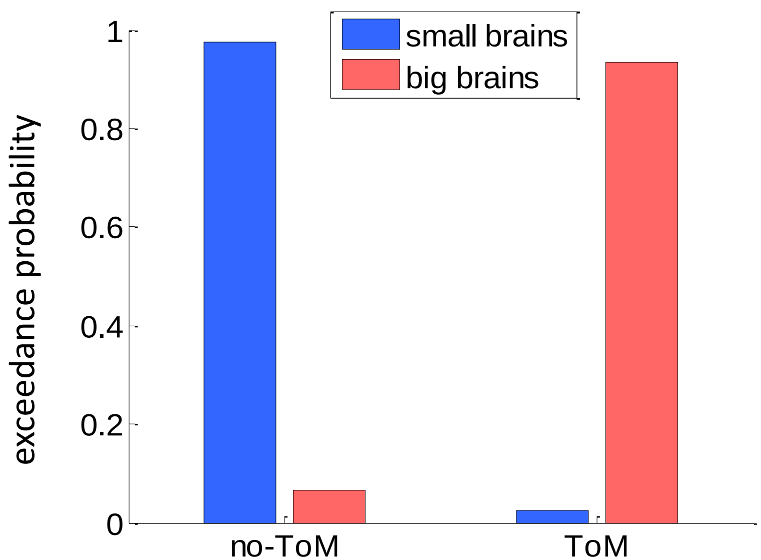
Assessing the *Machiavellian intelligence* hypothesis





Assessing the *social brain* hypothesis

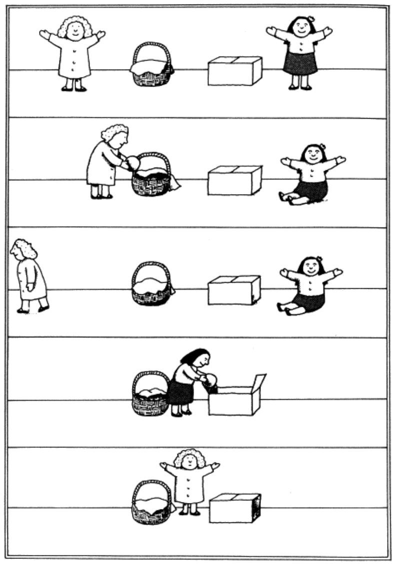


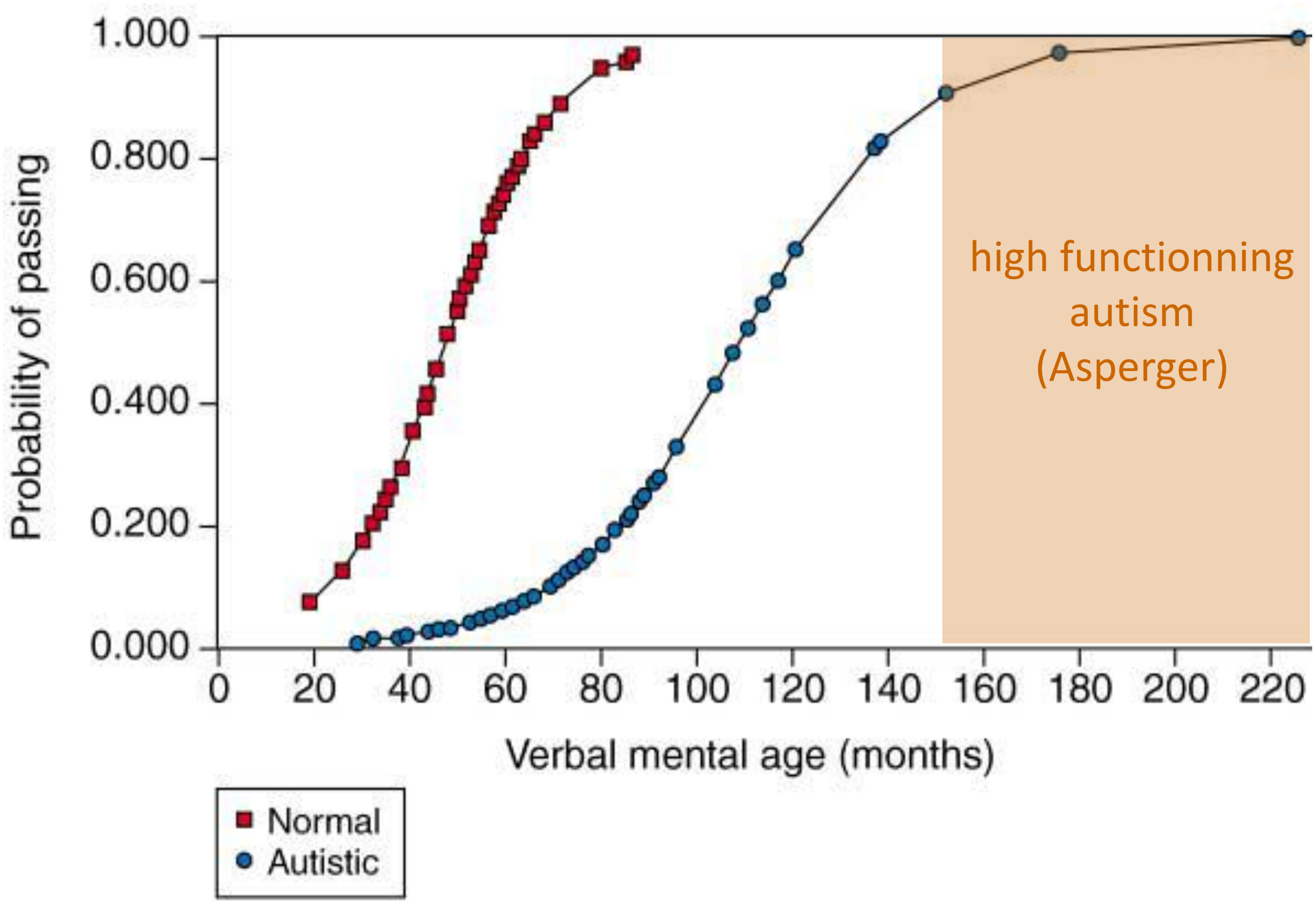


Overview of the talk

- ✓ Does ToM make a difference when we learn?
- ✓ Limited ToM sophistication: did evolution fool us?
- ✓ Playing *hide-and-seek* with non-human primates
- ✓ What about people with autism spectrum disorder?

ASD: ToM deficit hypothesis





[Baron-Cohen et al., 1985]

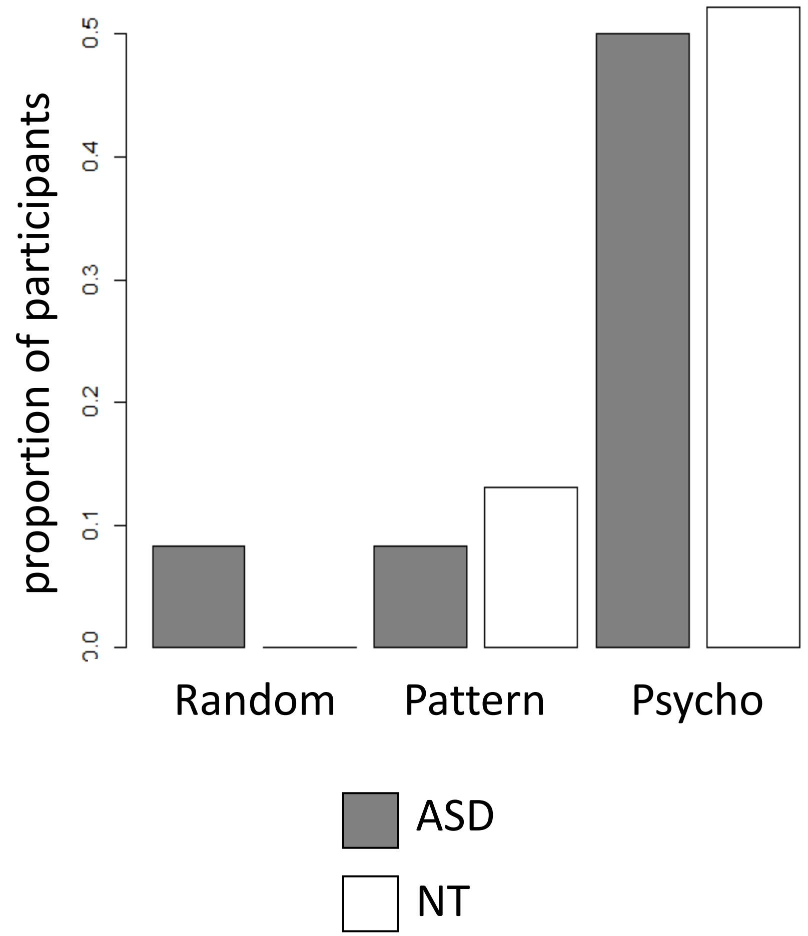
ASD patients: summary statistics

- High functioning autistic patients (N=24)
- Neurotypical participants (N=24)

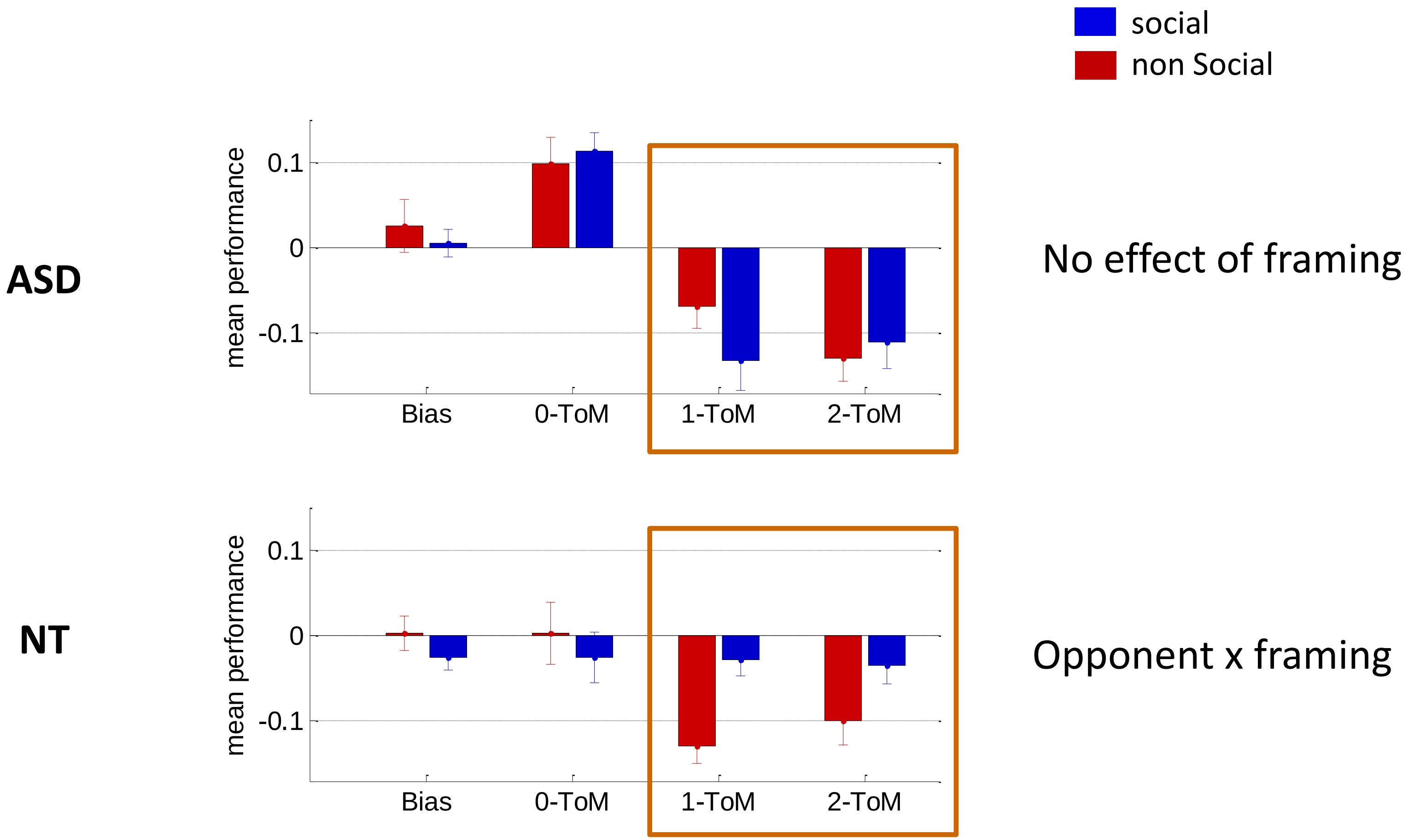
  matched for age, IQ, sex (21 males)

Group	ASD	NT
Age	25,5 (5,7)	27,9 (8,6)
IQ	104(17)	106 (14)
Social anhedonia	14,8 (8,4)	9.7 (4,2)

deceptive framing manipulation: sanity check

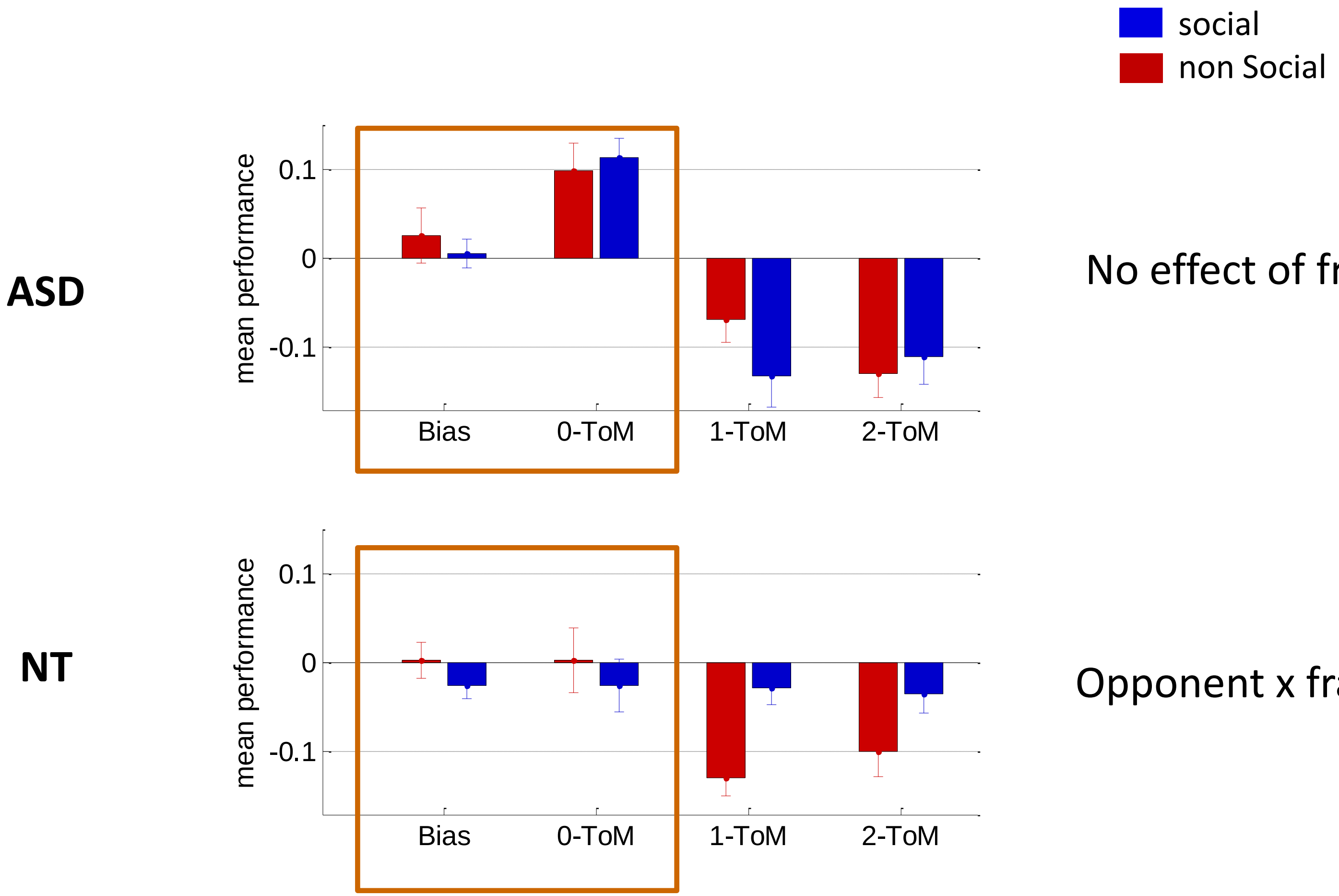


Behavioural performances: group comparison



Do autistic individuals have a deficit in recursive ToM? **Yes**

Behavioural performances: group comparison



No effect of framing

Opponent x framing

Do autistic individuals posses adaptive learning styles? **Yes**

Summary

- Meta-Bayesian inference
  - the brain's model of other brains assumes they are Bayesian too
  - reciprocal social interaction → recursive beliefs
- Does mentalizing make a difference when we learn?
  - social framing effect ("mentalize or be fooled")
  - distribution of ToM sophistication = mixed
- Evolution of ToM:
  - cooperation+learning → natural bounds to ToM sophistication
    ("being right is as good as being smart")
  - non-human primates → (brain) size matters
- Autism:
  - ASD = 1-ToM ? (cannot consider that others are mentalizing too)

References and acknowledgements

This is **Marie Devaine**'s hard work!

I also would like to thank:

- Dr. **G. Hollard**, economist (Ecole Polytechnique, Palaiseau, France)
- Dr. **S. Masi**, ethologist (Natural History Museum, Paris, France)
- Dr. **B. Forgeot d'Arc**, psychiatrist (Hopital Rivière-Des-Prairies, Montréal, Canada)

The social Bayesian brain: does mentalizing matter when we learn?
M. Devaine, G. Hollard, J. Daunizeau
PLoS Comp. Biol. (2014), 10(12): e1003992.

Theory of Mind: did evolution fool us?
M. Devaine, G. Hollard, J. Daunizeau
PLoS ONE (2014), 9(2): e87619.

Observing the observer (II): deciding when to decide
J. Daunizeau, H. E. M. Den Ouden, M. Pessiglione, S. J. Kiebel, K. J. Friston, K. E. Stephan
PLoS ONE (2010b), 5(12): e15555.

Observing the observer (I): meta-Bayesian models of learning and decision-making
J. Daunizeau, H. E. M. Den Ouden, M. Pessiglione, K. E. Stephan, S. J. Kiebel, K. J. Friston
PLoS ONE (2010a), 5(12): e15554.