On the adaptive fitness of the social sense: lessons from Bayesian Decision Theory

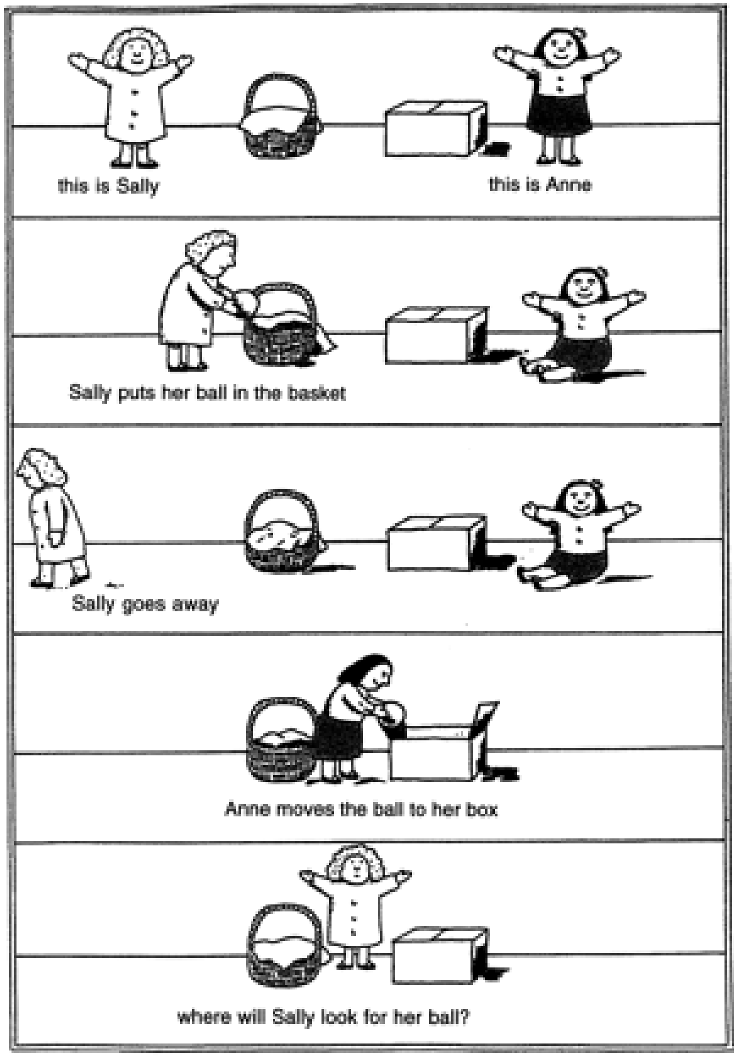
Dr. Marie Devaine

Jean Daunizeau

Institut du Cerveau et de la Moelle épinière (ICM) – Paris, France

ETH – Zurich, Switzerland

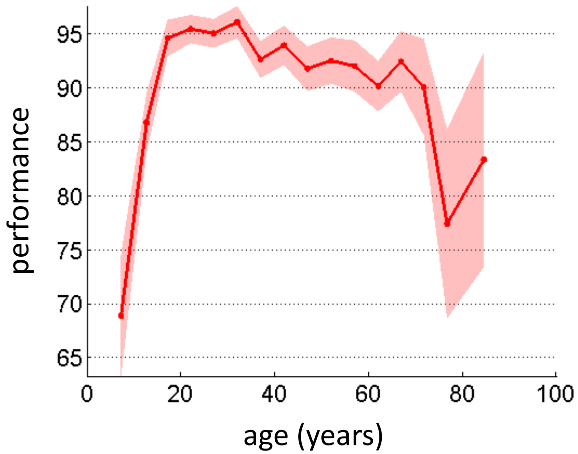
Theory of Mind: the "false belief" test



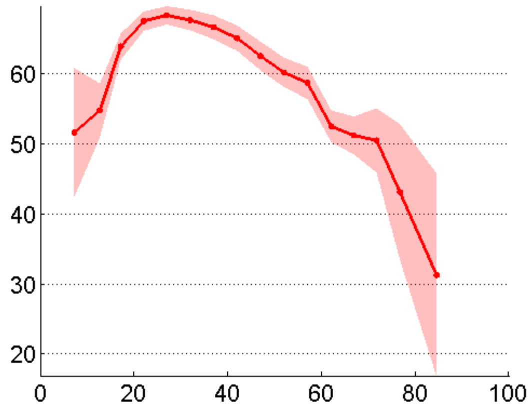
Wimmer & Perner, 1983

Lifespan dynamics of Theory of Mind

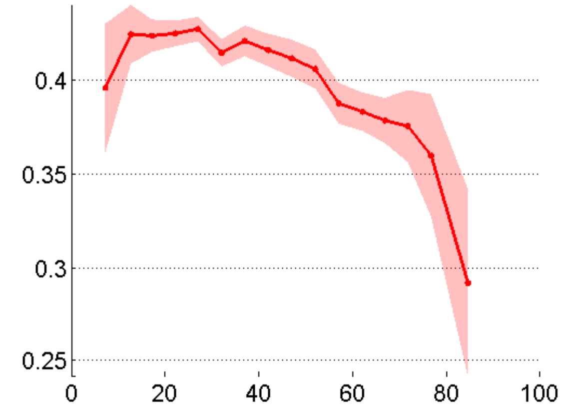
false belief test
(n = 8884)



Frith-Happé animations
(n = 6098)

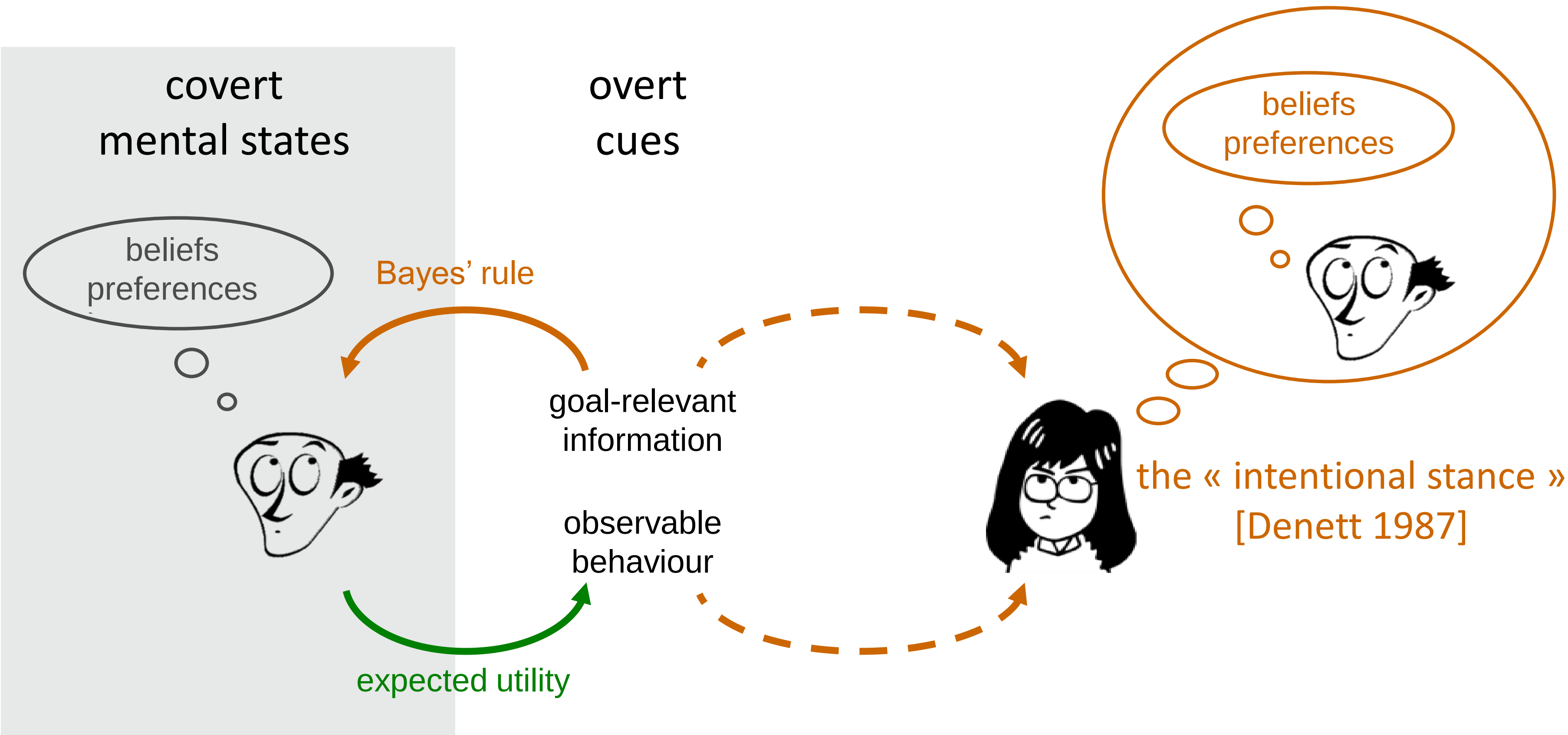


hide-and-seek
(n = 5926)



Klindt et al., 2016

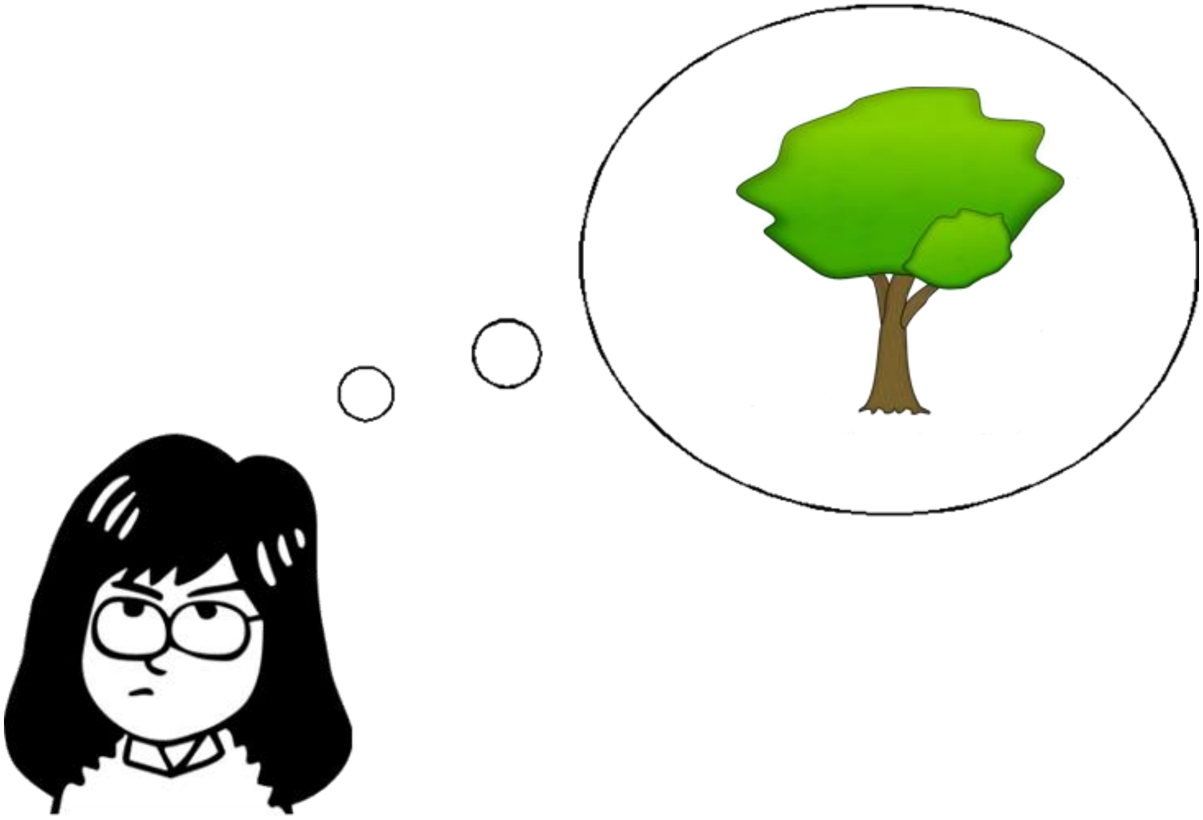
What computational problem does ToM solve?



1) ToM = *inverse* Bayesian Decision Theory?

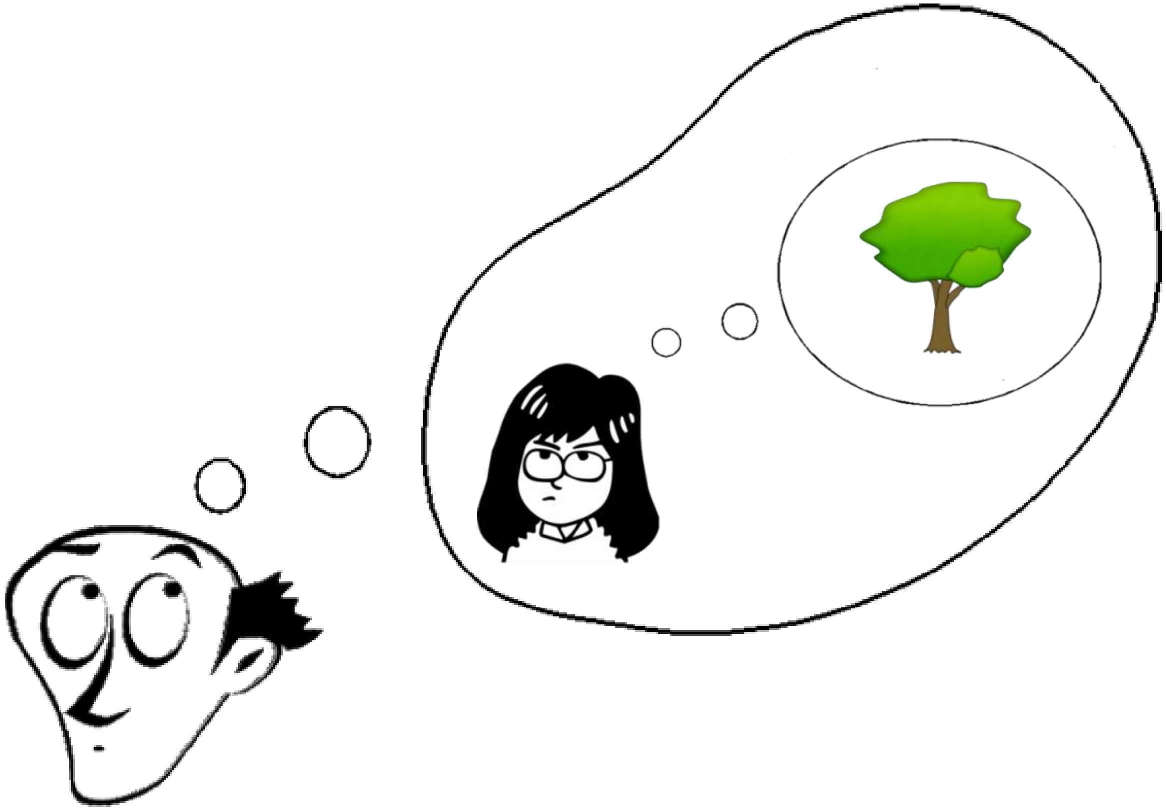
2) ToM sophistication = depth of recursive beliefs?

playing without ToM (*0-ToM*)



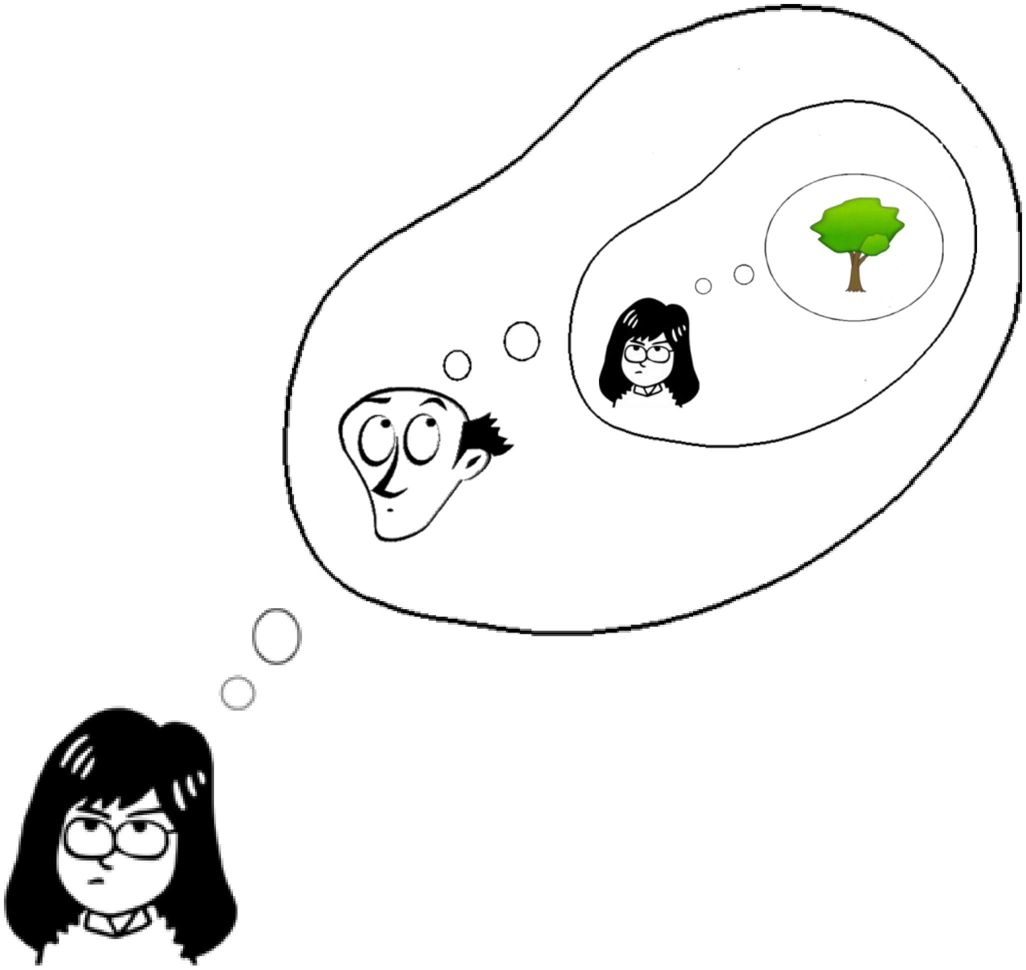
« I believe that you will hide behind the tree »

playing with ToM: recursive beliefs (*1-ToM*)



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

playing with ToM: recursive beliefs (*2-ToM*)



« I believe that you believe that I believe … »

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?
- ✓ A short lesson from fMRI

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?
- ✓ A short lesson from fMRI

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:

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

- Nulling the ensuing Free-Energy derivative yields the *[k+1]-ToM* learning rule:

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

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

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

k-ToM: VB belief update rule

$$p_t^{op} = \sum_{l<\kappa} \lambda_t^{k,\kappa} \; p_t^{op,\kappa}$$

$$p_t^{op,\kappa} \approx s \circ \tilde{v}^{\kappa}\left(\mu_{t-1}^{k,\kappa}, \Sigma_{t-1}^{k,\kappa}\right)$$

$$\lambda_t^{k,\kappa} \approx \left[\frac{\lambda_{t-1}^{k,\kappa} \; p_t^{op,\kappa}}{\sum_{\kappa'<k} \lambda_{t-1}^{k,\kappa'} \; p_t^{op,\kappa'}}\right]^{a_t^{op}} \left[\frac{\lambda_{t-1}^{k,\kappa}\left(1-p_t^{op,\kappa}\right)}{\sum_{\kappa'<k} \lambda_{t-1}^{k,\kappa'}\left(1-p_t^{op,\kappa'}\right)}\right]^{1-a_t^{op}}$$

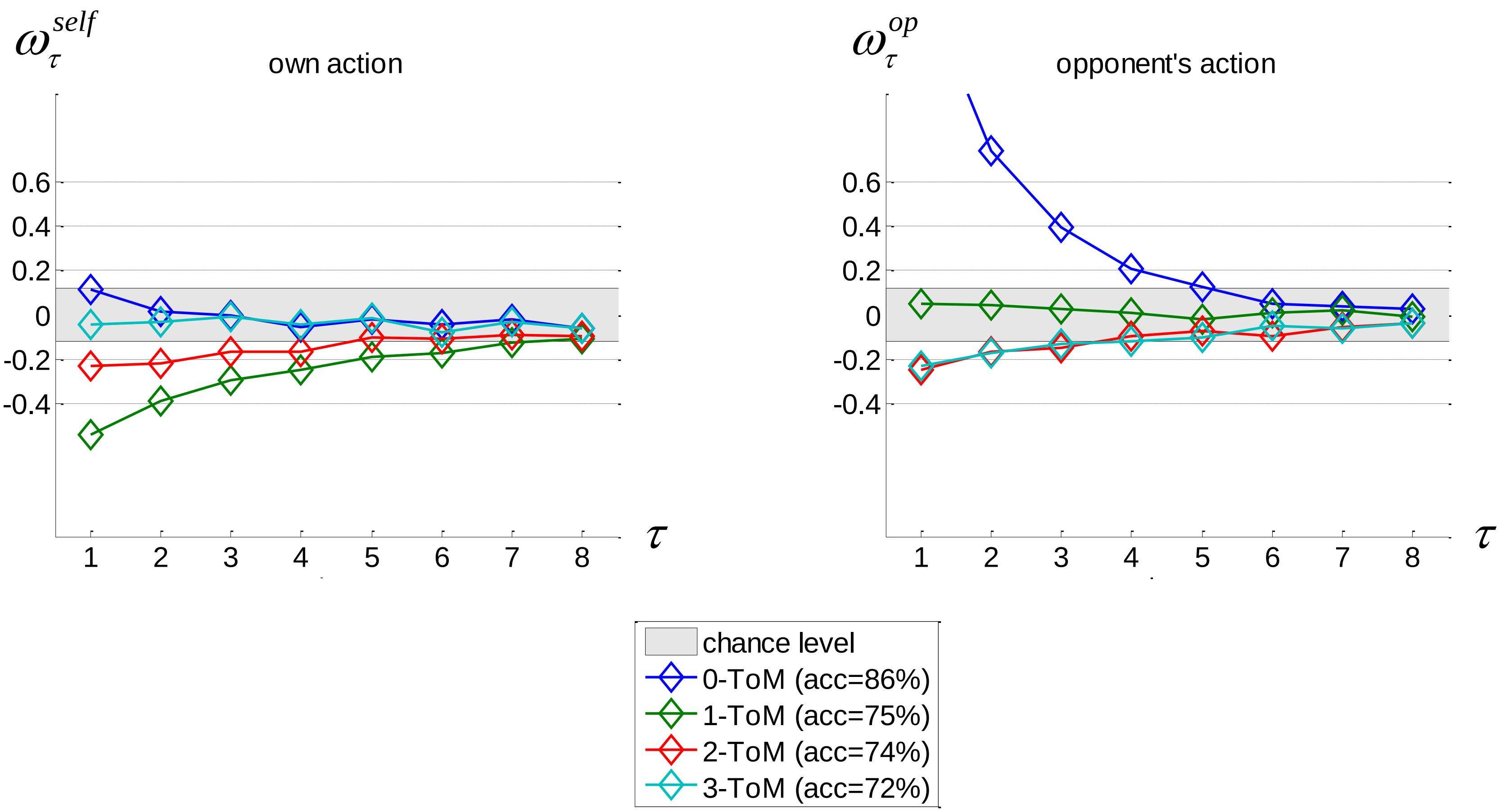
$$\mu_t^{k,\kappa} \approx \mu_{t-1}^{k,\kappa} + \lambda_t^{\kappa} \; \Sigma_t^{k,\kappa} \; W_{t-1}^{\kappa}\left(a_t^{op} - s \circ v^{\kappa}\left(\mu_{t-1}^{k,\kappa}\right)\right)$$

$$\Sigma_t^{k,\kappa} \approx \left[\left(\Sigma_{t-1}^{k,\kappa} + \sigma^{k}\right)^{-1} + s' \circ v^{\kappa}\left(\mu_{t-1}^{k,\kappa}\right) \lambda_t^{\kappa} \; W_{t-1}^{\kappa\,T} W_{t-1}^{\kappa}\right]^{-1}$$

$$v^1\left(x_t^1\right) = \frac{p_t^{self}\,\Delta U^1 + \left(1-p_t^{self}\right)\Delta U^0}{\beta_t}$$

k-ToM's learning rule in competitive games

1st-order Volterra decomposition: $$p\left(a_t^{self}=1 \middle| \omega\right) = s\left(\omega^0 + \sum_\tau \omega_\tau^{op}\ a_{t-\tau}^{op} + \sum_\tau \omega_\tau^{self}\ a_{t-\tau}^{self}\right)$$



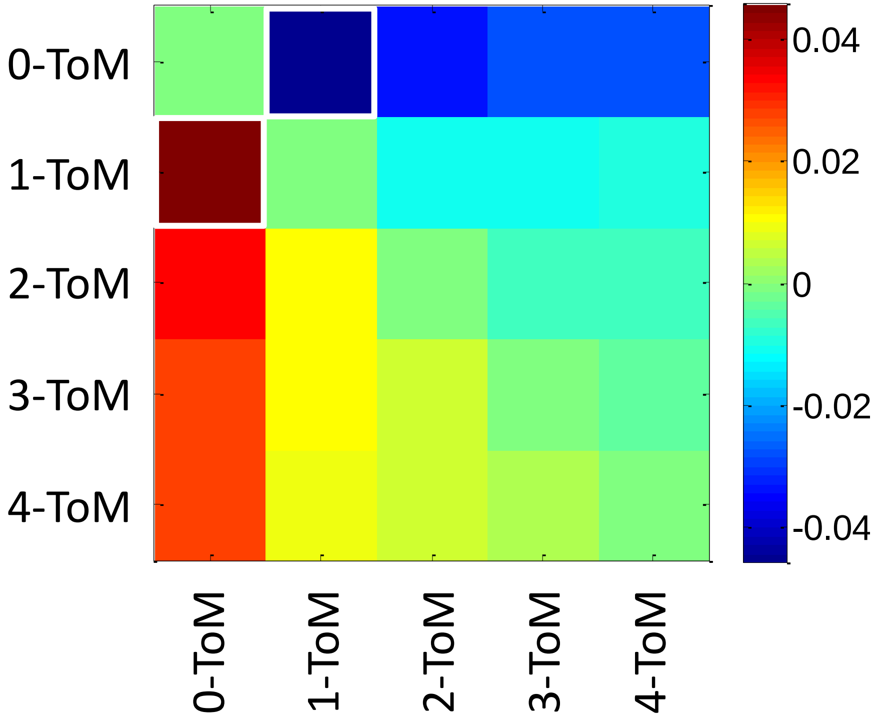
k-ToM's performance in competitive games

payoff 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

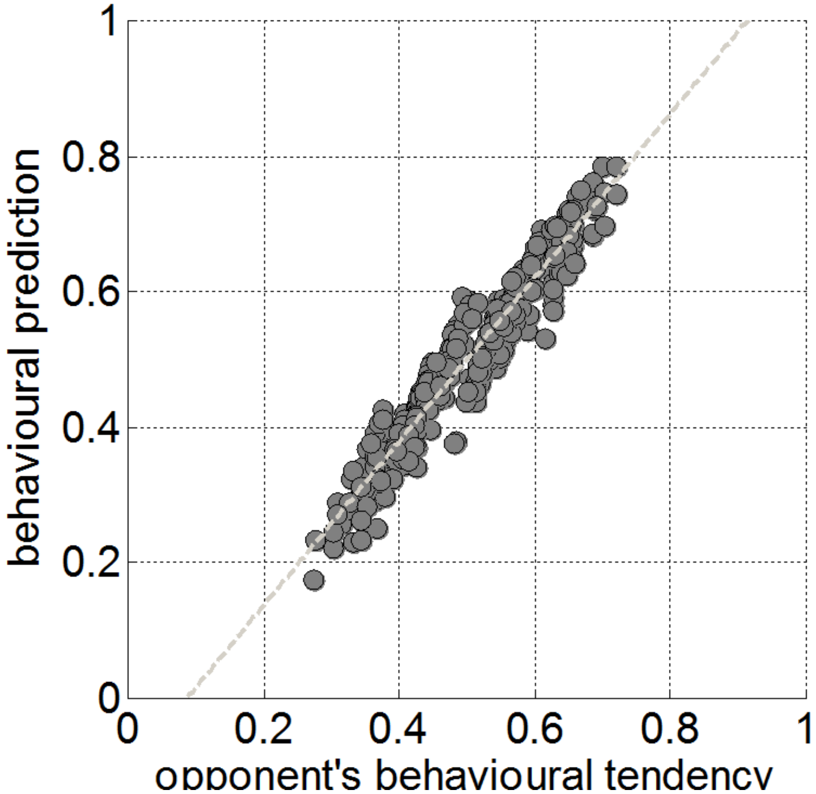
simulated
behavioural performance
(#wins/trial)

$\tau = 512$

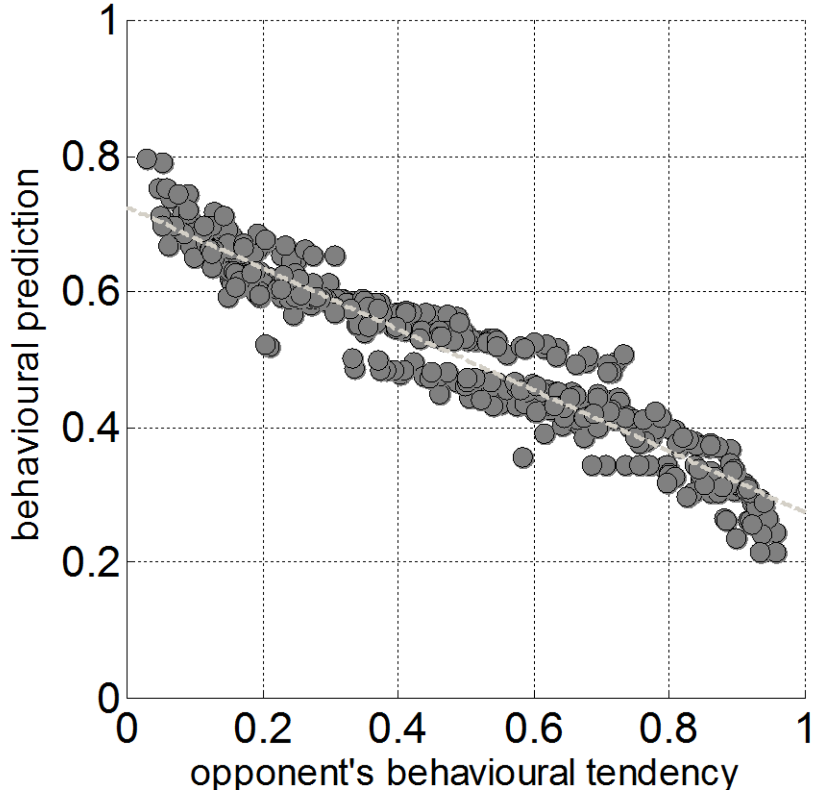


Everybody is somebody's fool

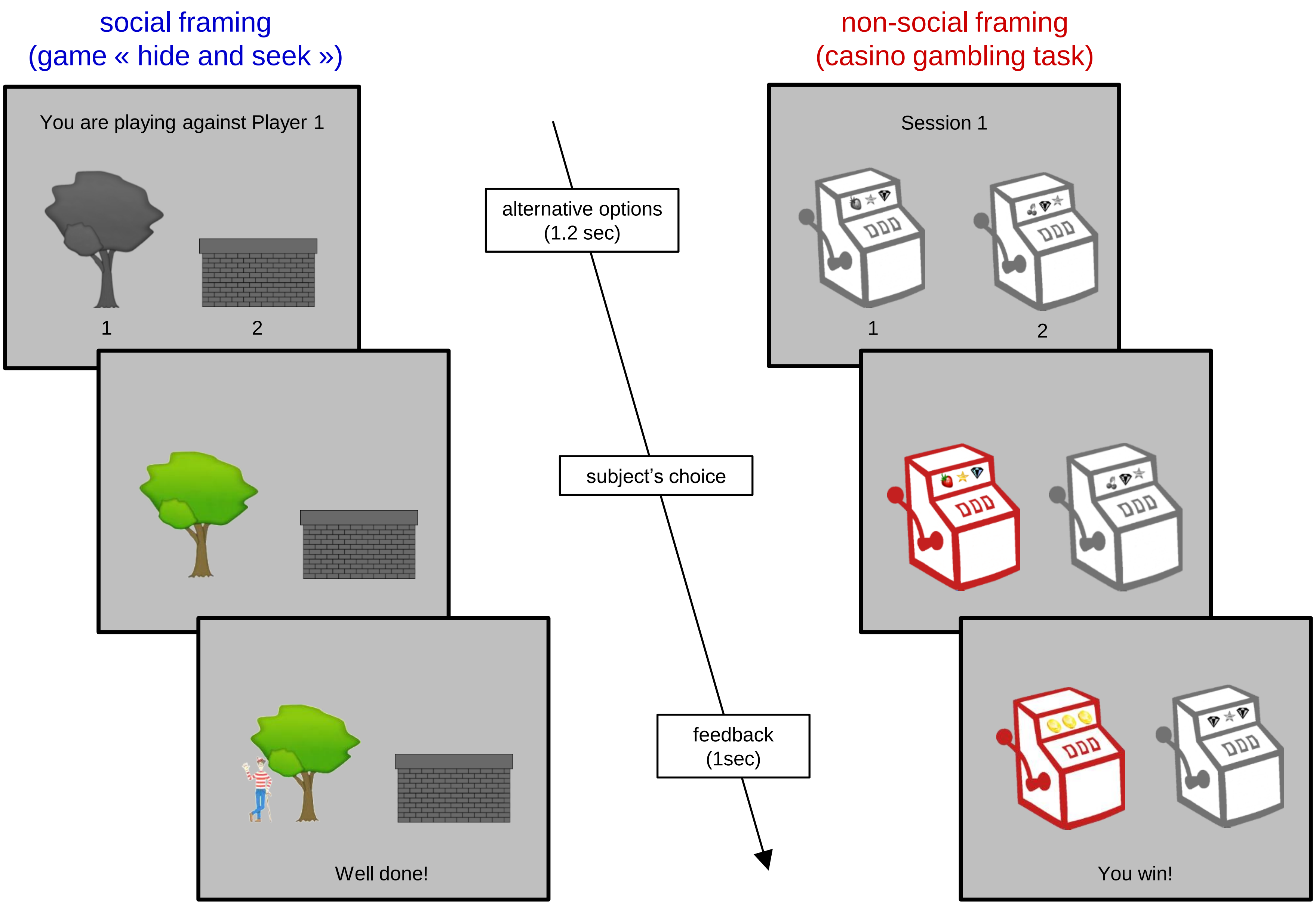
1-ToM predicts 0-ToM



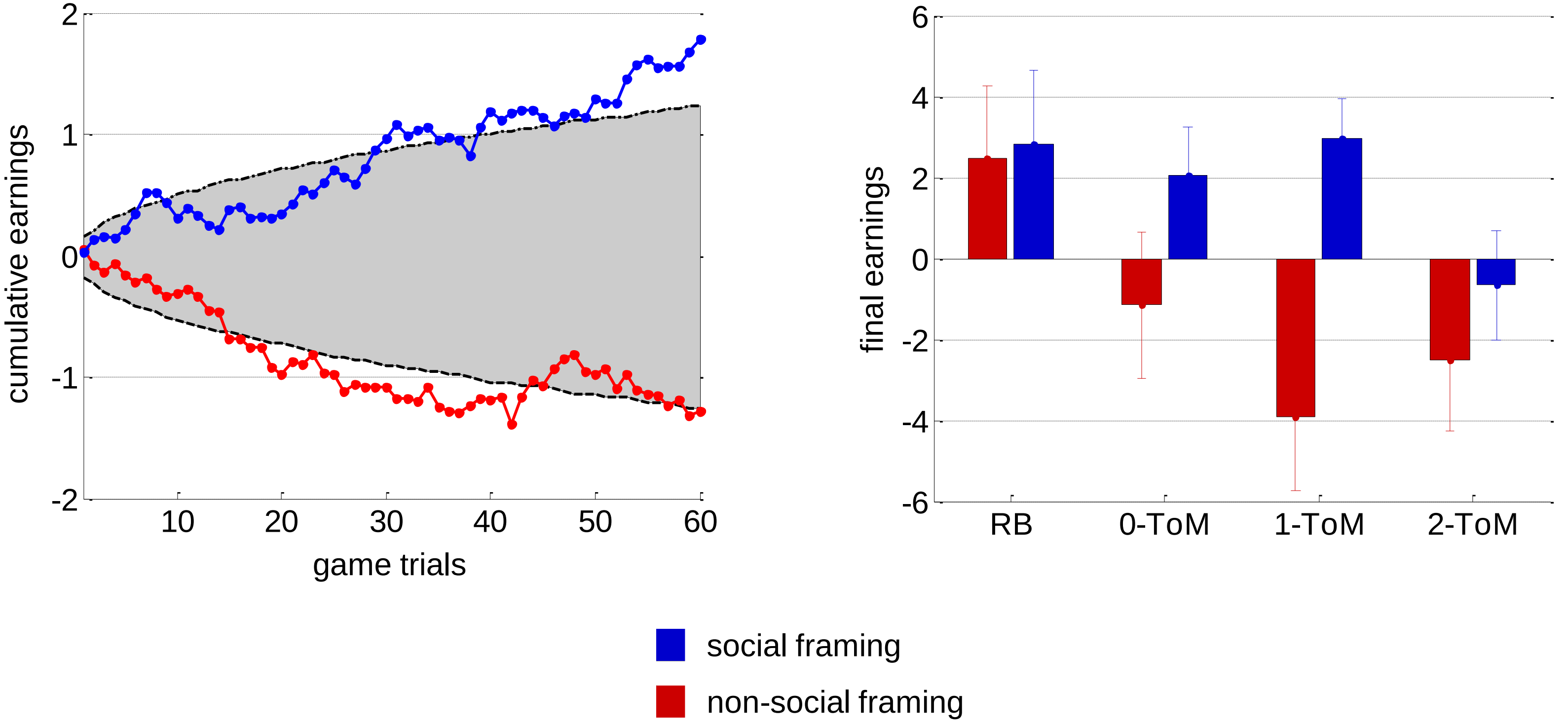
0-ToM predicts 1-ToM



Behavioural task design

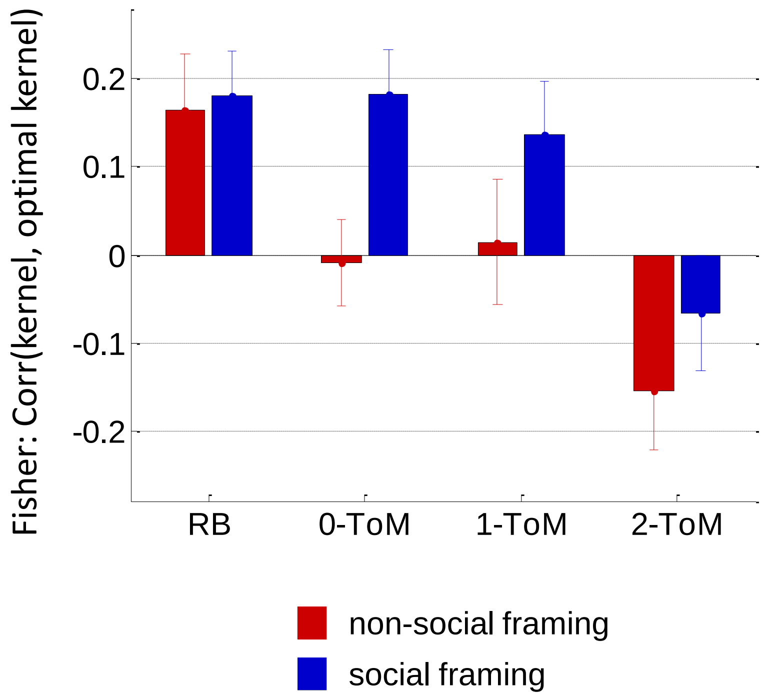


Behavioural performances (N=26)



Devaine et al. (2014a)

Learning style: similarity to best *k-ToM* response



ANOVA:

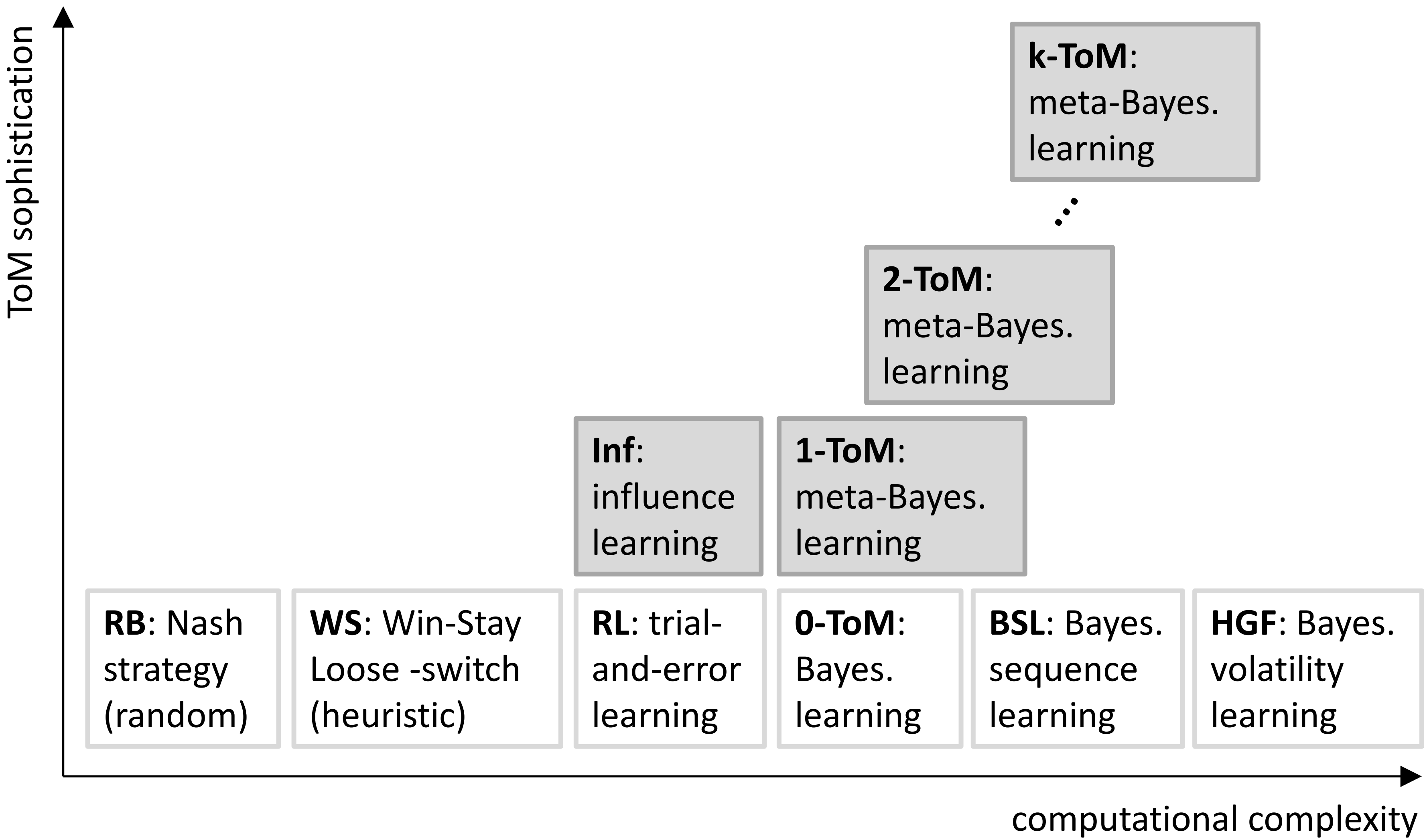
- **Framing** (p=0.02), **op** (p=0.0001), 0 framingXop
- 0 age, 0 sex

SOBEL:

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

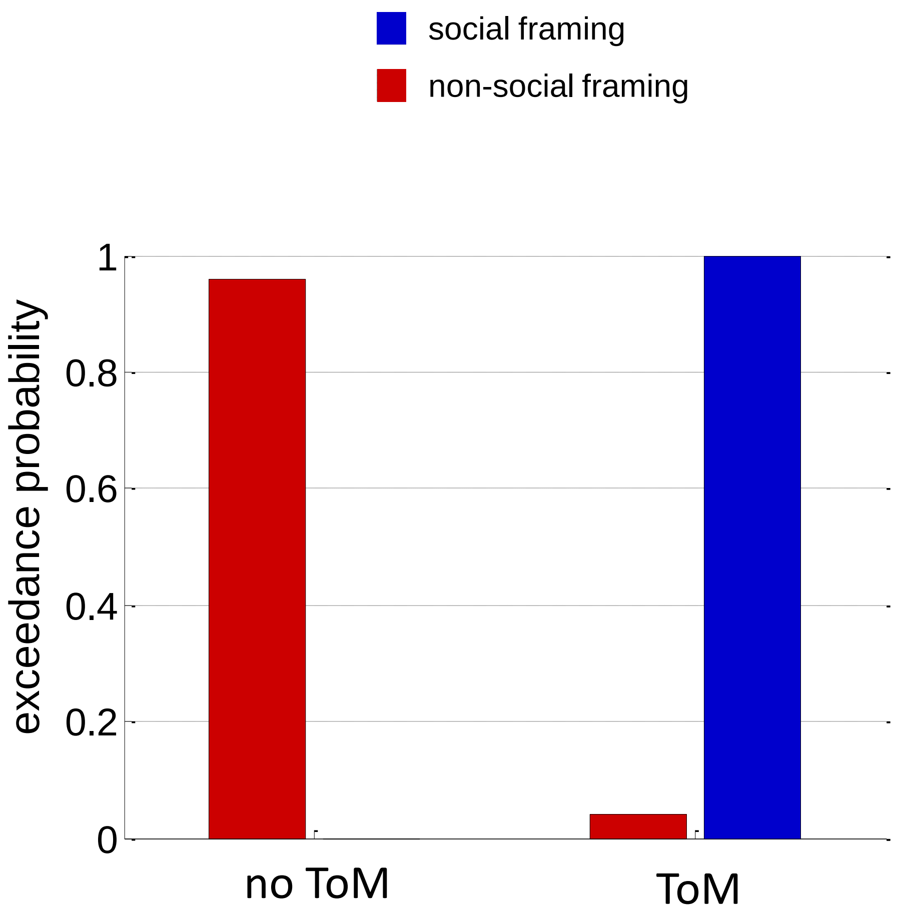
Devaine et al. (2014a)

Trial-by-trial choice sequences: learning *styles*



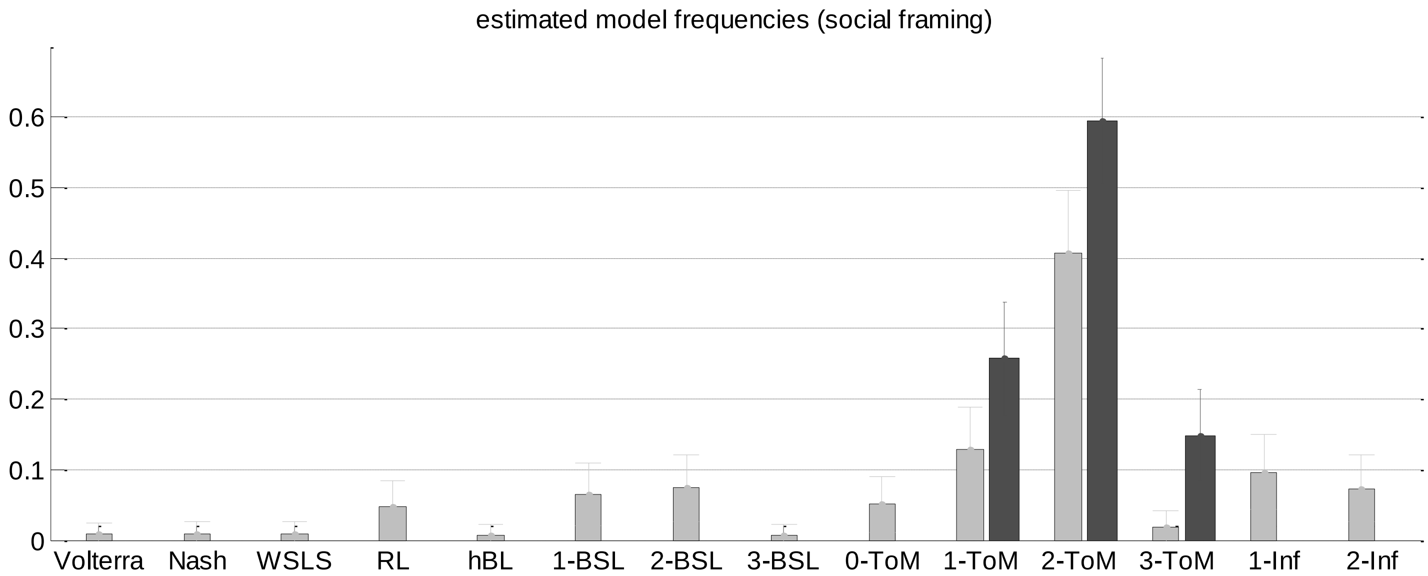
Devaine et al. (2014a)

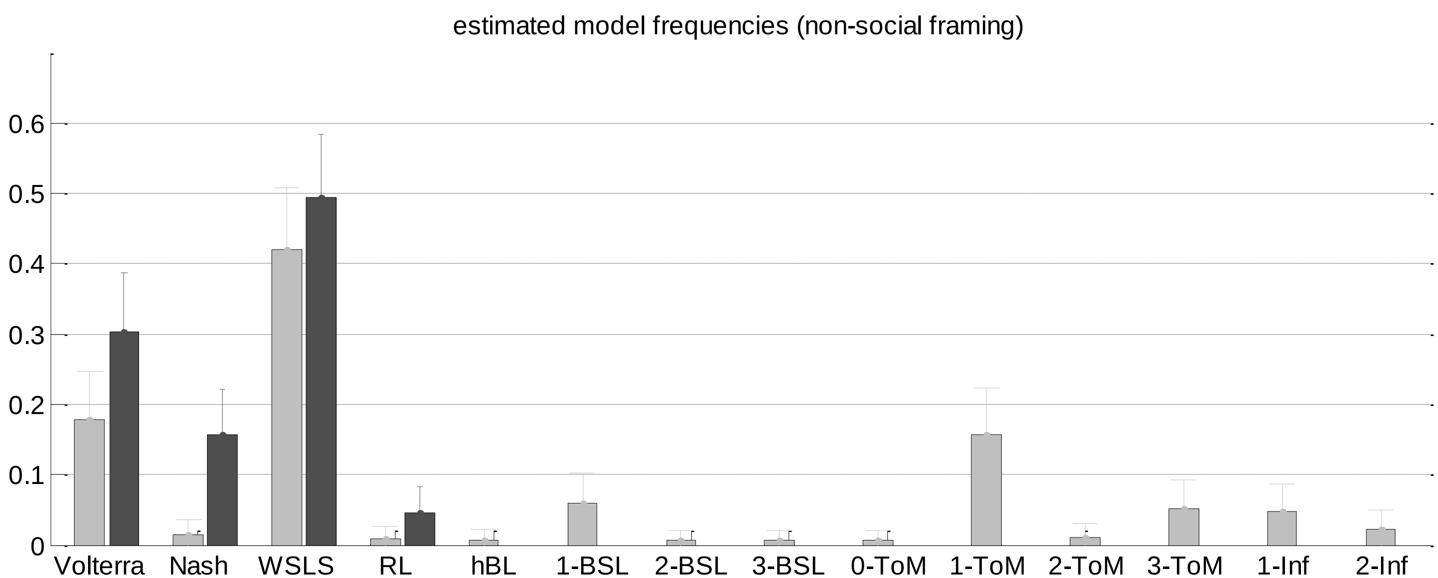
Analysis of trial-by-trial choice sequences



Devaine et al. (2014a)

Variability of human ToM sophistication





Devaine et al. (2014a)

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?
- ✓ A short lesson from fMRI

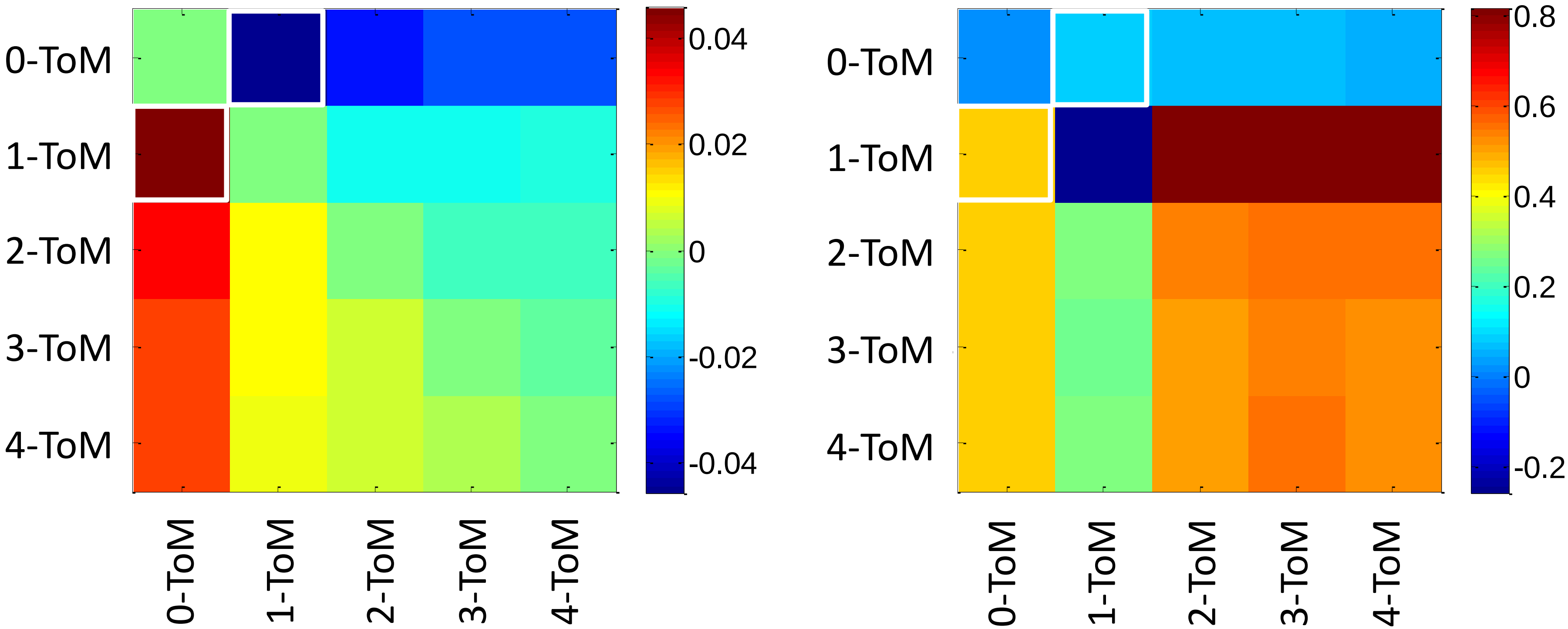
Competitive versus cooperative games

« 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

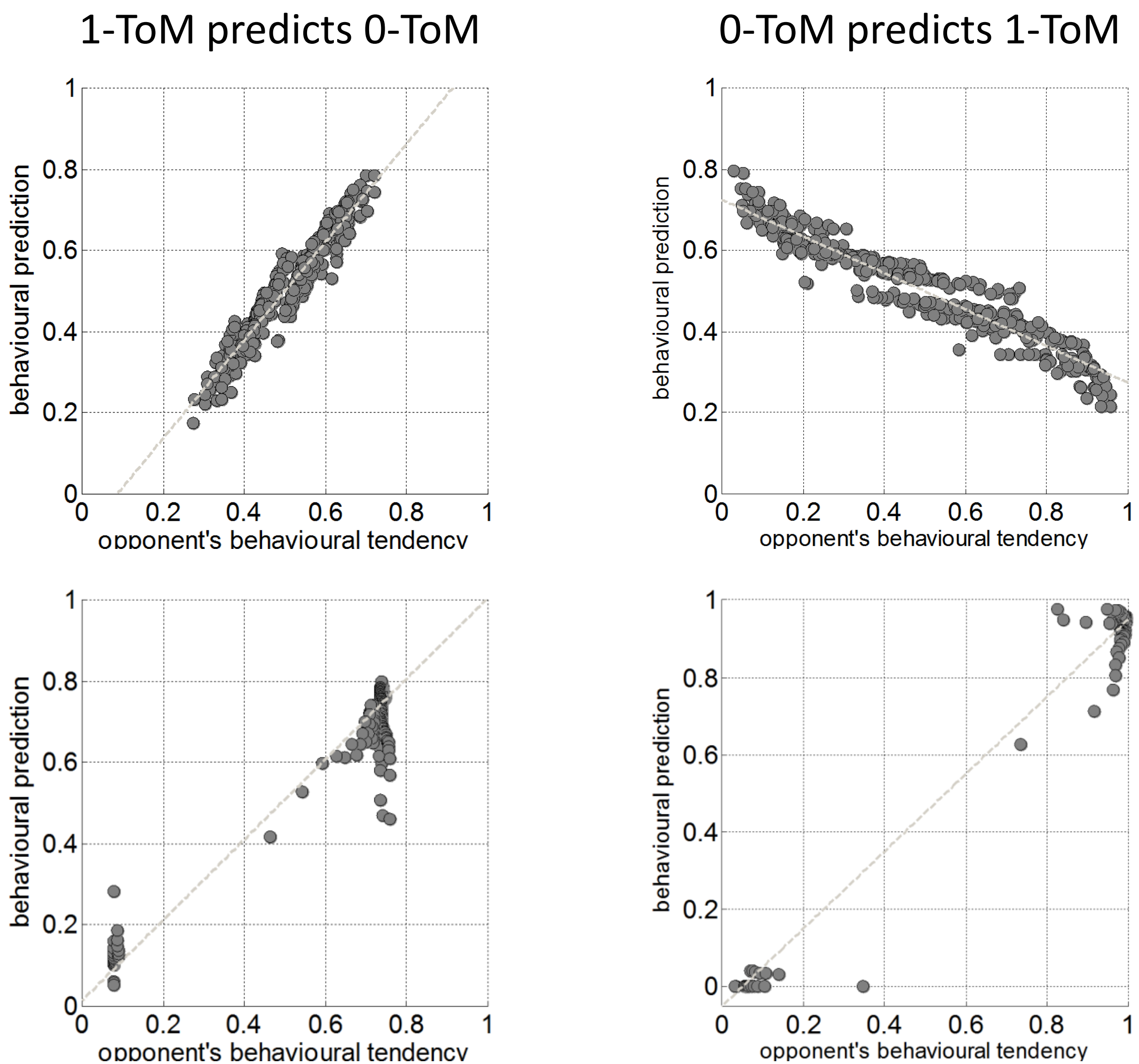


Devaine et al. (2014b)

Being right is as good as being smart

« hide and seek »

« battle of the sexes »



Devaine et al. (2014b)

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

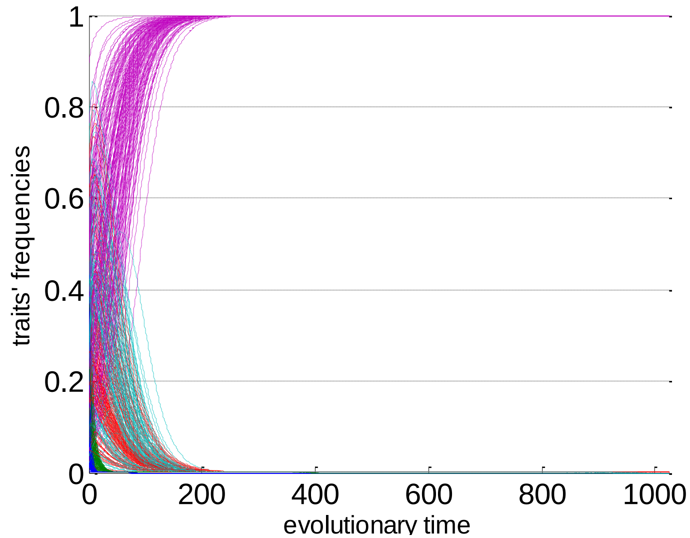
→ Replicator dynamics [Maynard-Smith 1982, Hofbauer 1998]:

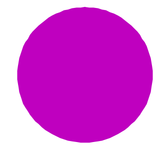
$$\frac{ds}{dt} = Diag(s)\left(\sum_i \omega_i \, Q^{(i)}(\tau) s - \sum_i \omega_i \, s^T Q^{(i)}(\tau) s\right)$$

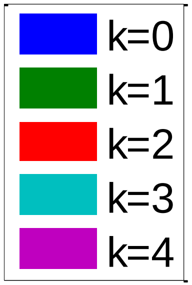
evolutionary stable states: $s_\infty \equiv \lim_{t\to\infty} s(t)$

Replicator dynamics and ESS

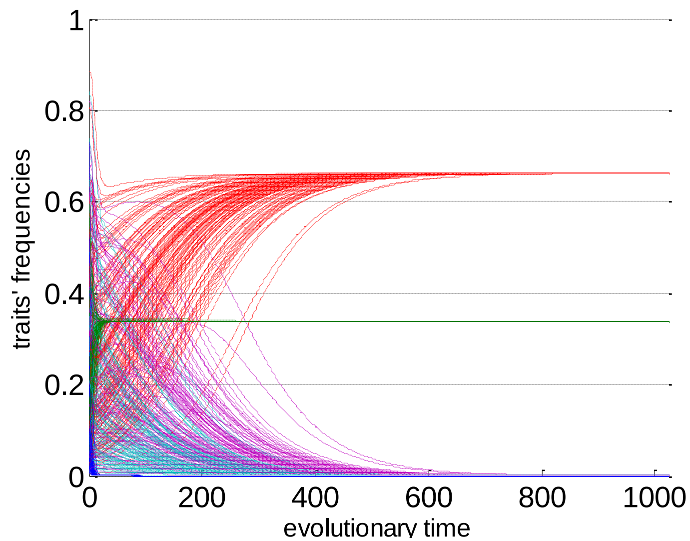
« hide and seek »

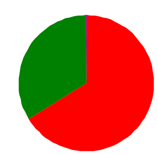






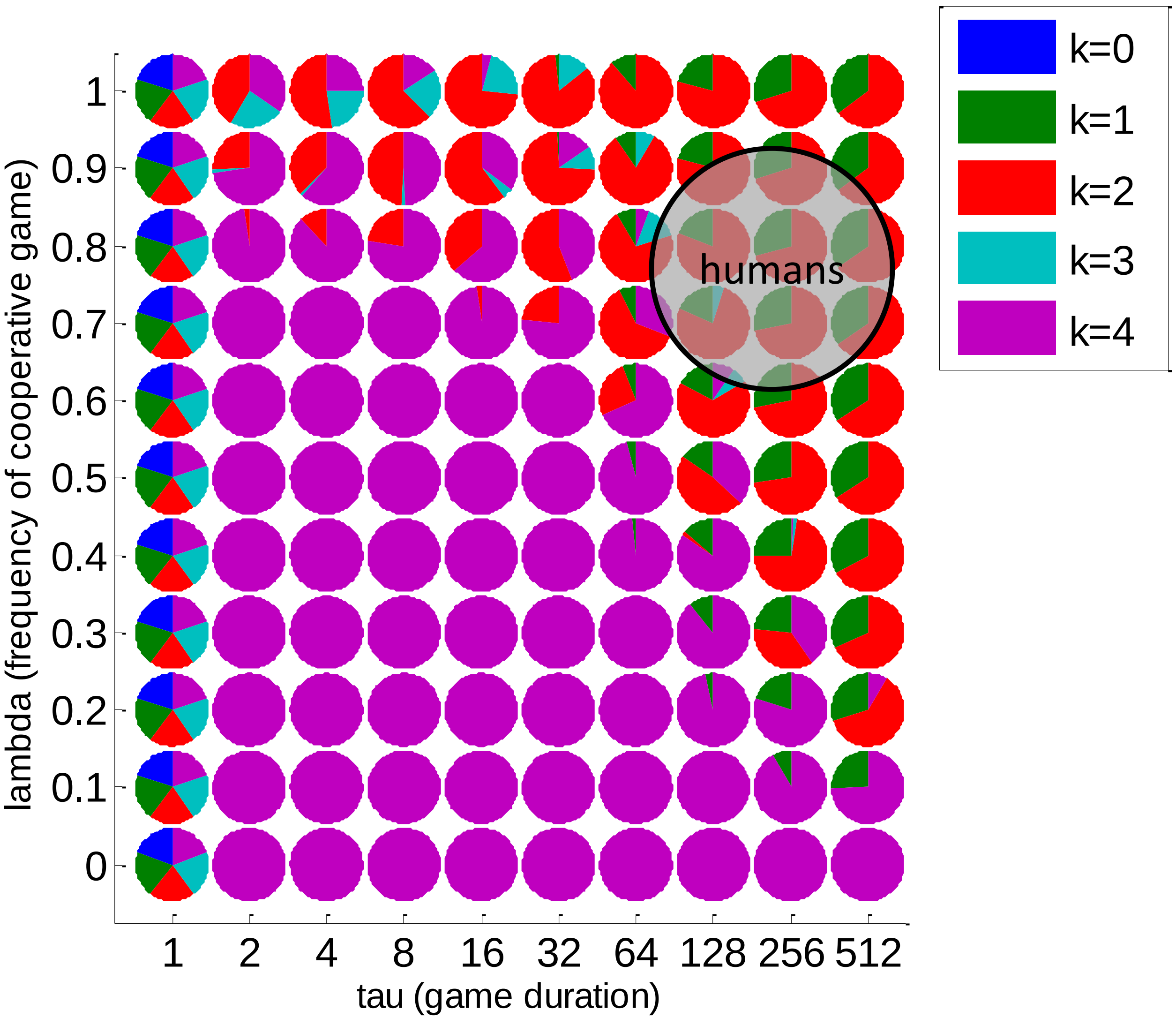
« battle of the sexes »





Devaine et al. (2014b)

ESS: phase portrait

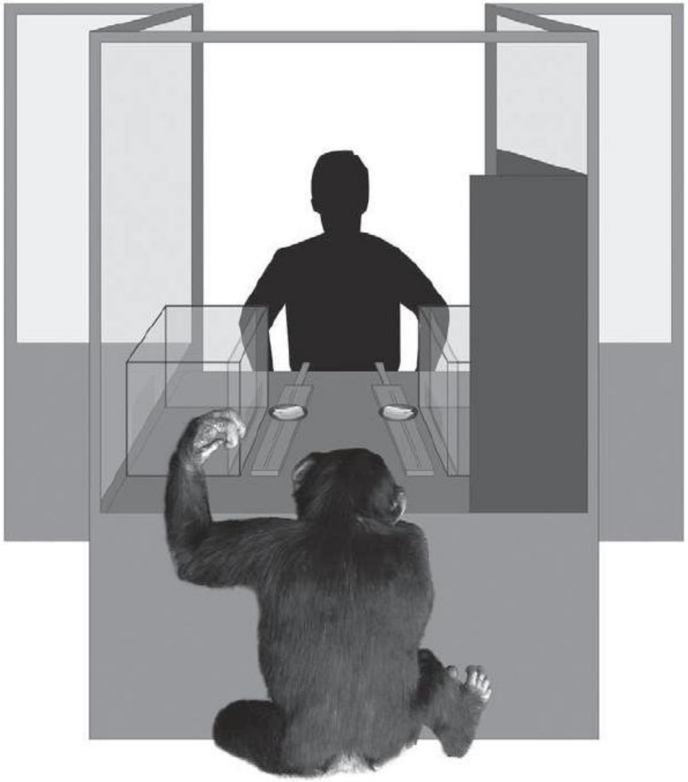


Devaine et al. (2014b)

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?
- ✓ A short lesson from fMRI

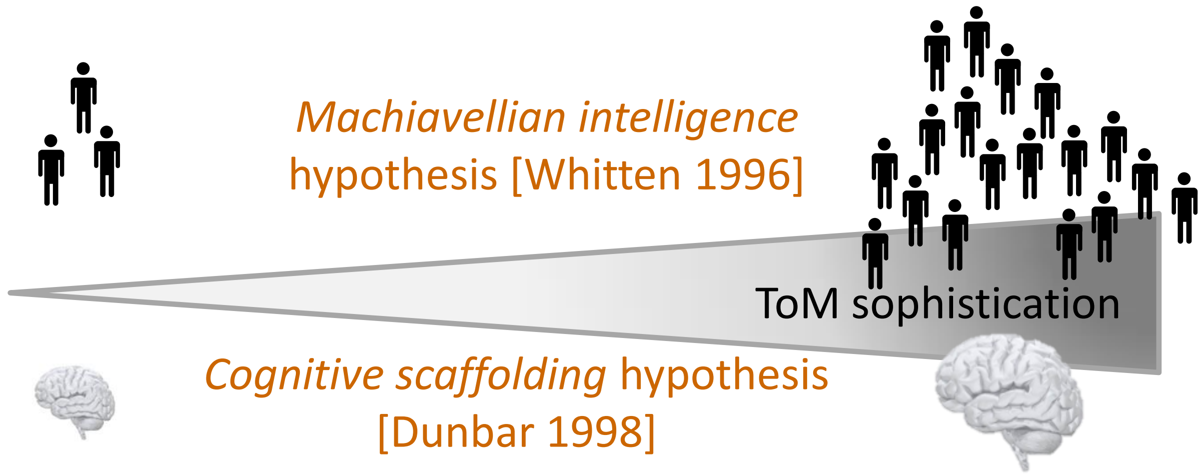
The main confound in primates' ToM assessment

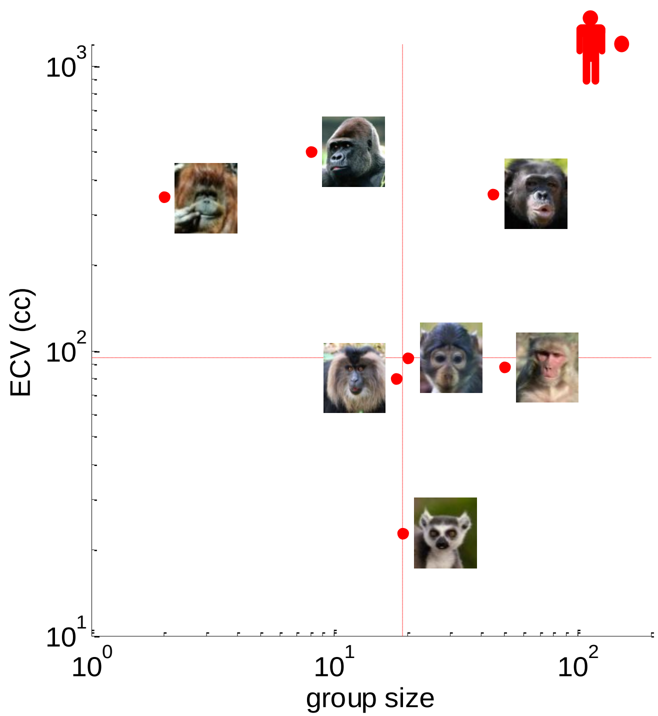


You're competing for the food.
Where should you approach the food from?

[Hare 2006]

Evolutionary pressure on ToM sophistication





r=-0.28, p=0.54

Devaine et al. (submitted)

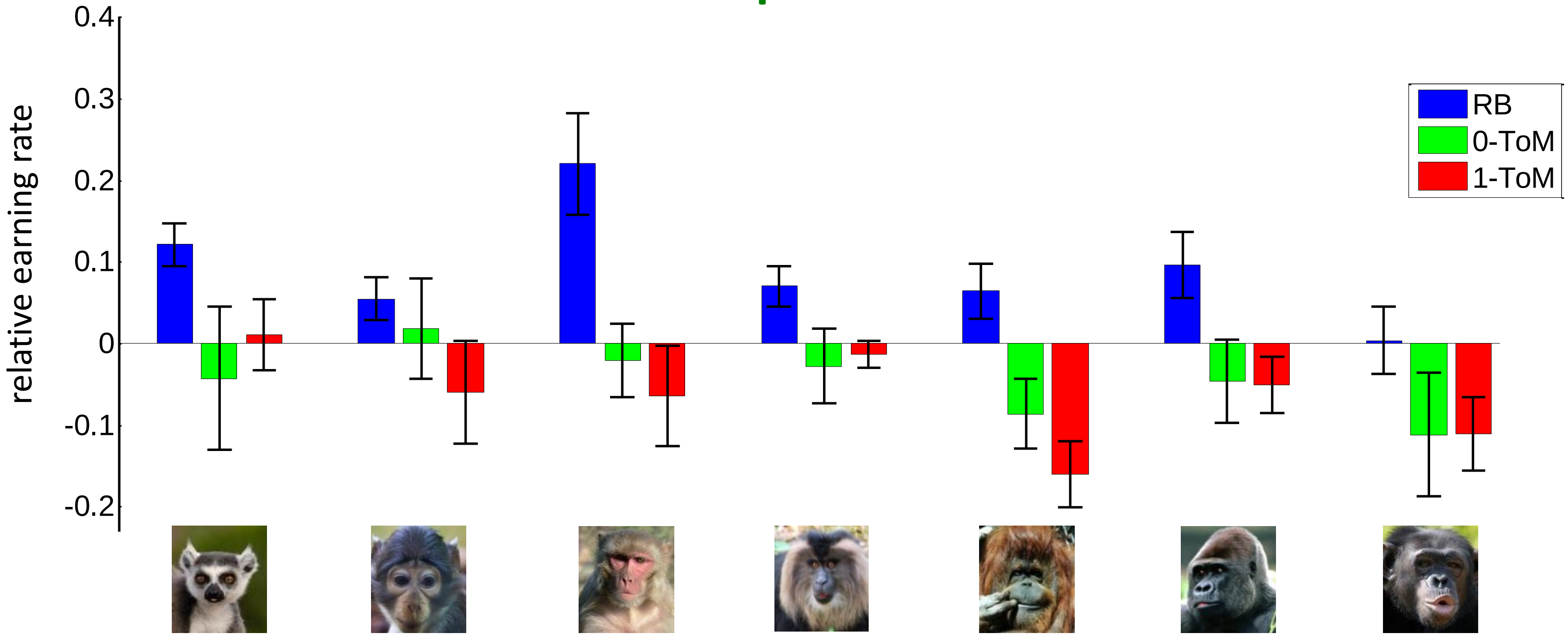
Playing “hide and seek“ with primates

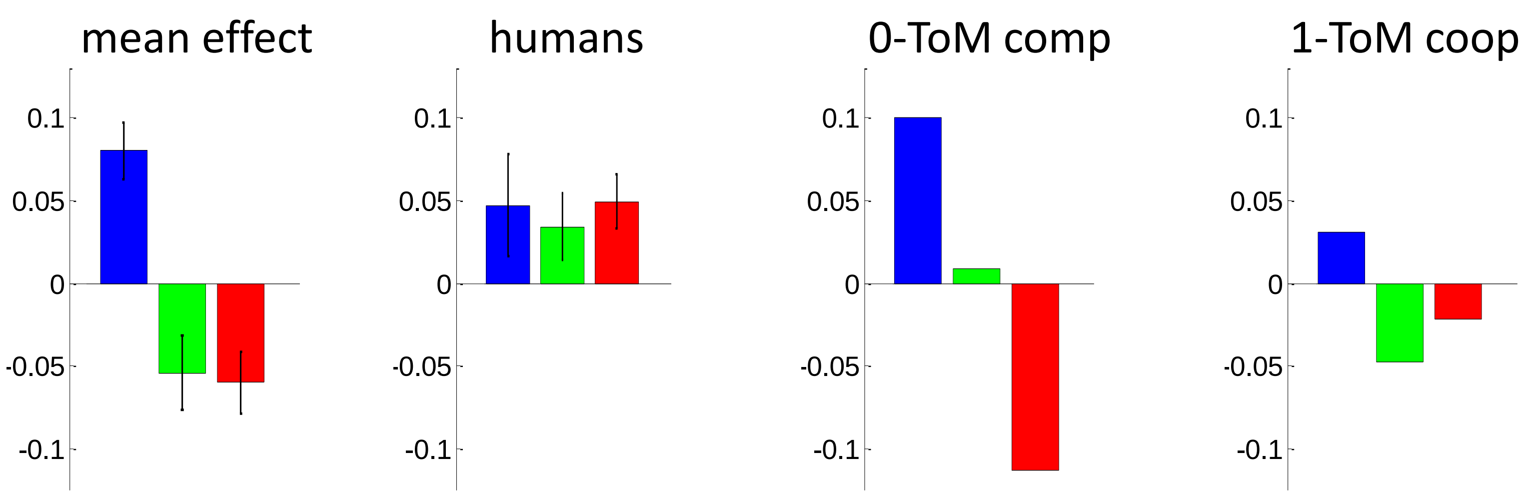
• ***Subjects (n=39):***
Macaques (4+5), Orangutans (7), Chimps (6), Gorillas (5), Mangabeys (8), Lemurs (4)

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

Devaine et al. (submitted)

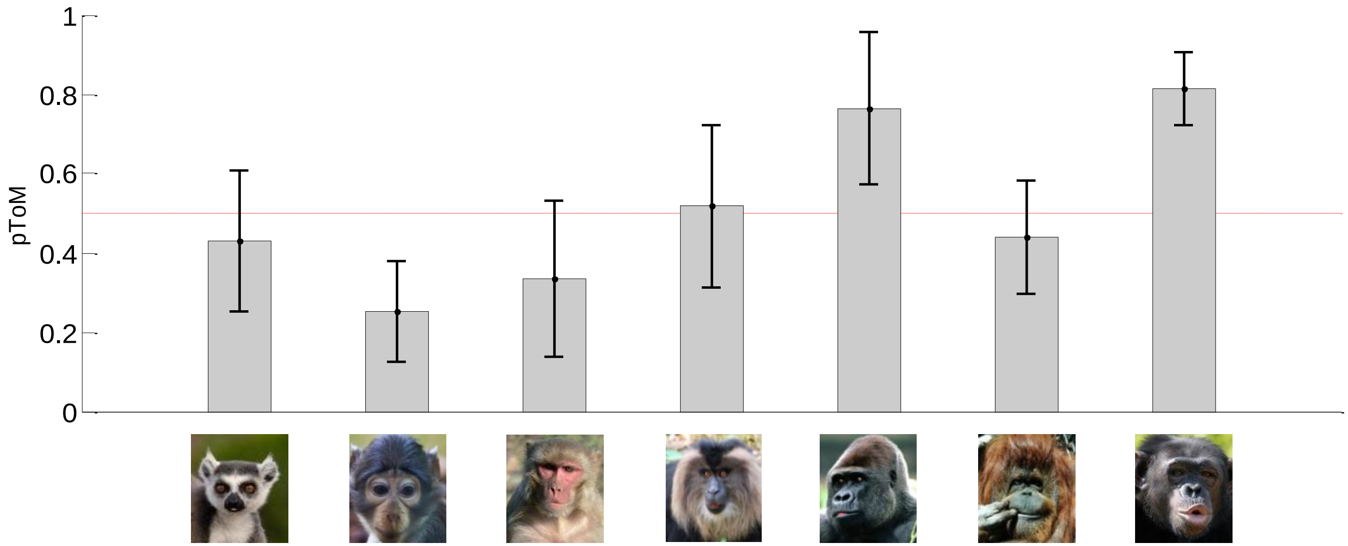
Behavioural performances

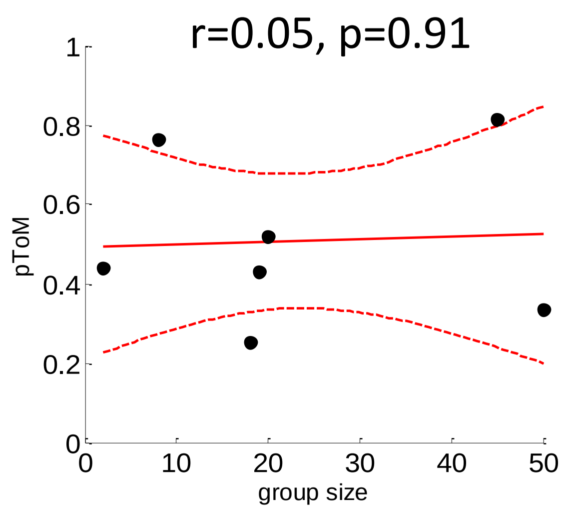


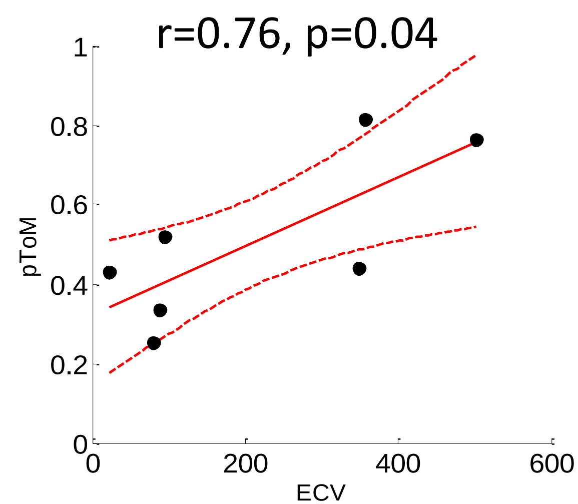


Devaine et al. (submitted)

(brain) size matters

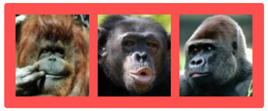




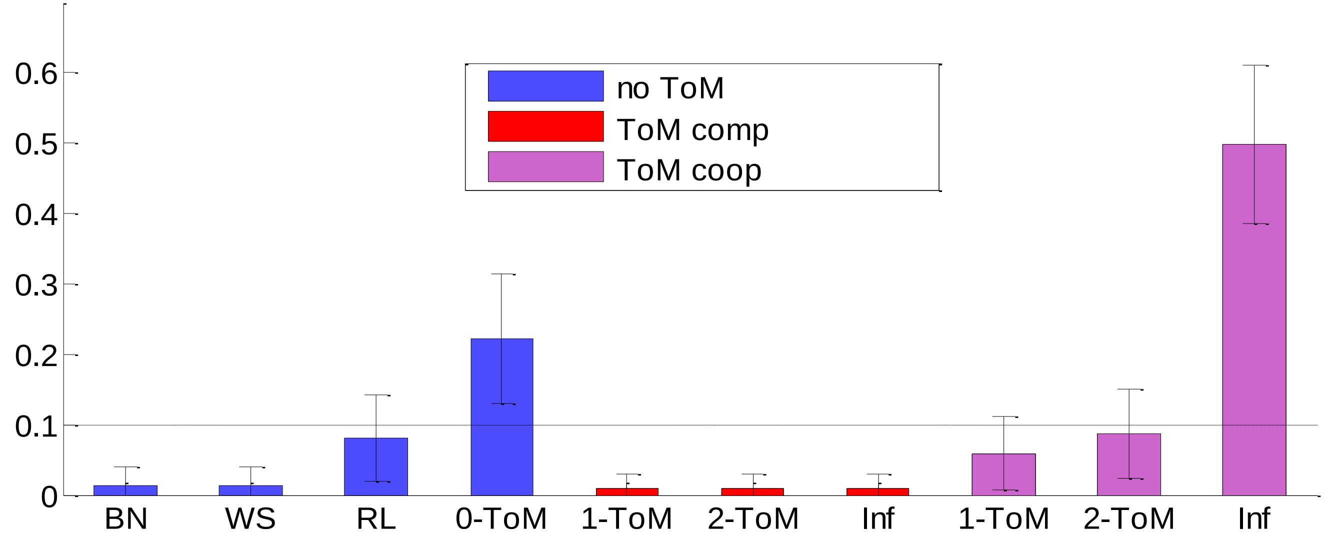


Devaine et al. (submitted)

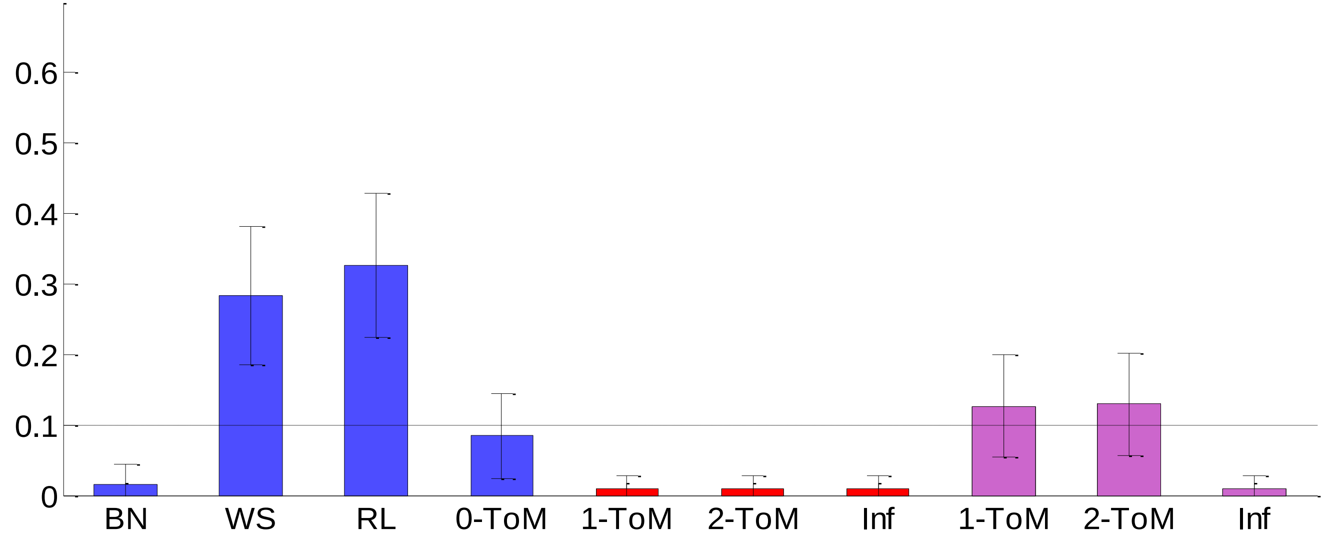
Variability of non-human ToM sophistication



large brains



small brains

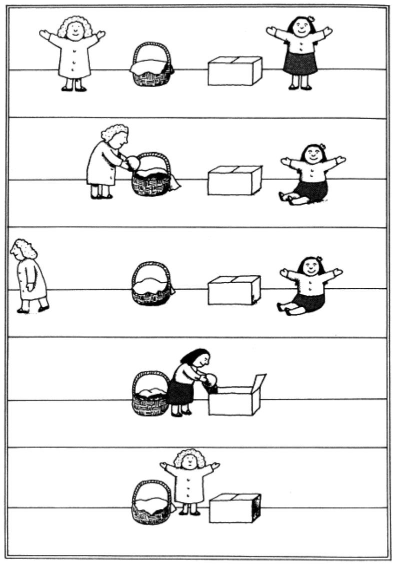


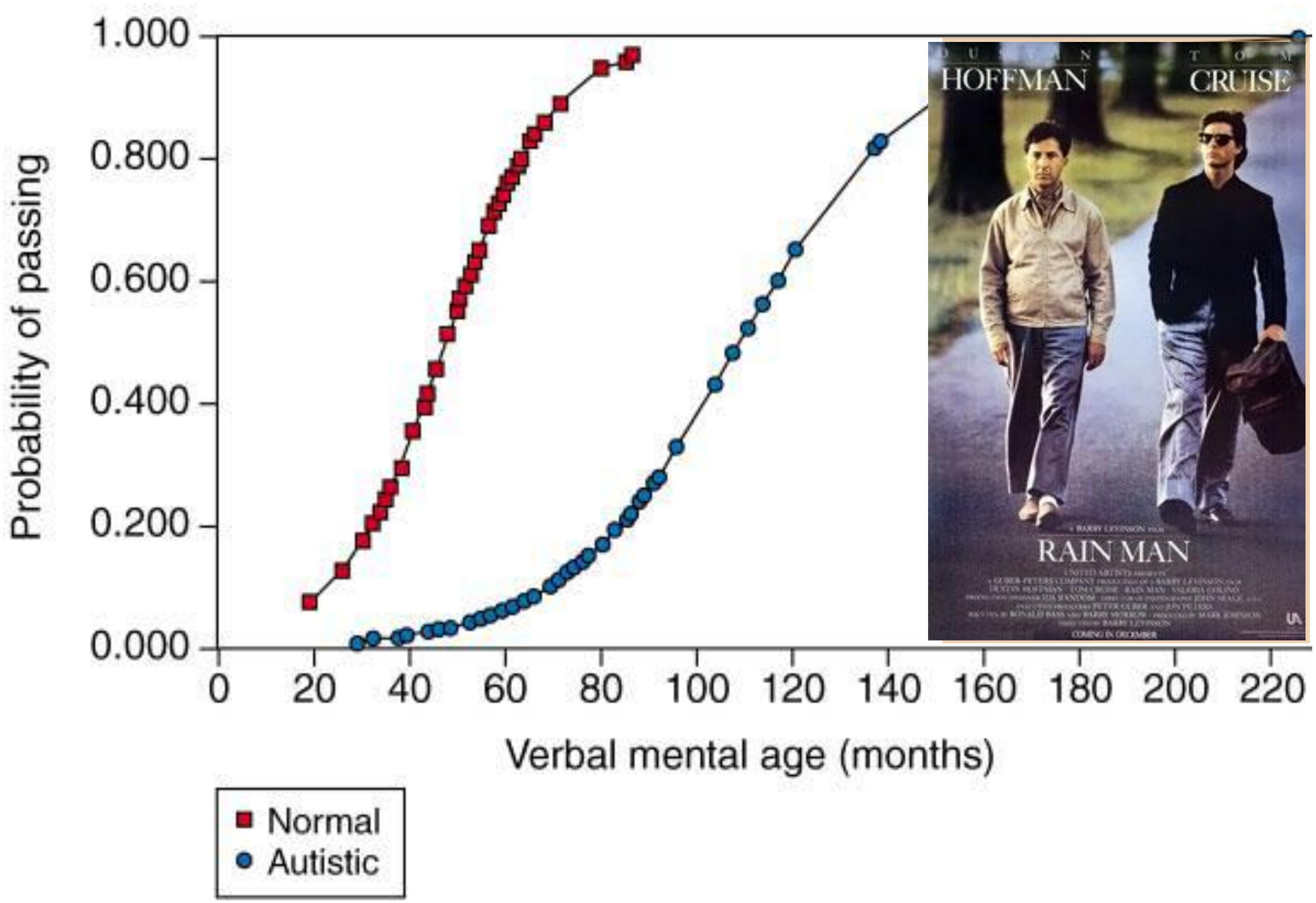
Devaine et al. (submitted)

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?
- ✓ A short lesson from fMRI

ASD: ToM deficit hypothesis





[Baron-Cohen et al., 1985]

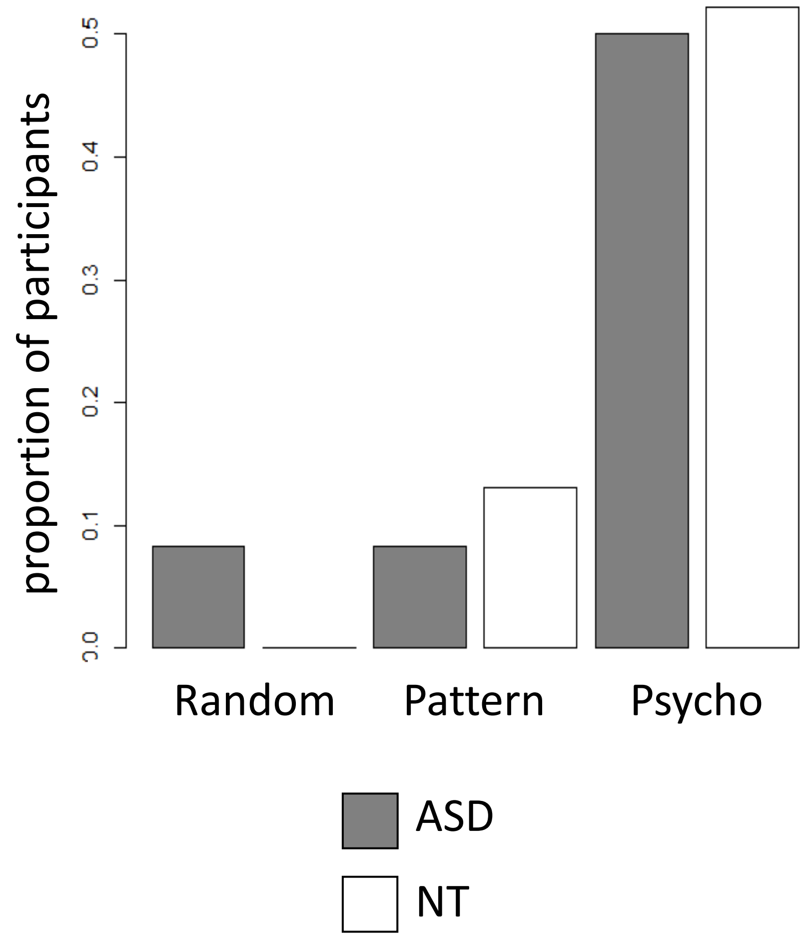
ASD patients: summary statistics

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

  matched for age, IQ, sex (21 males)

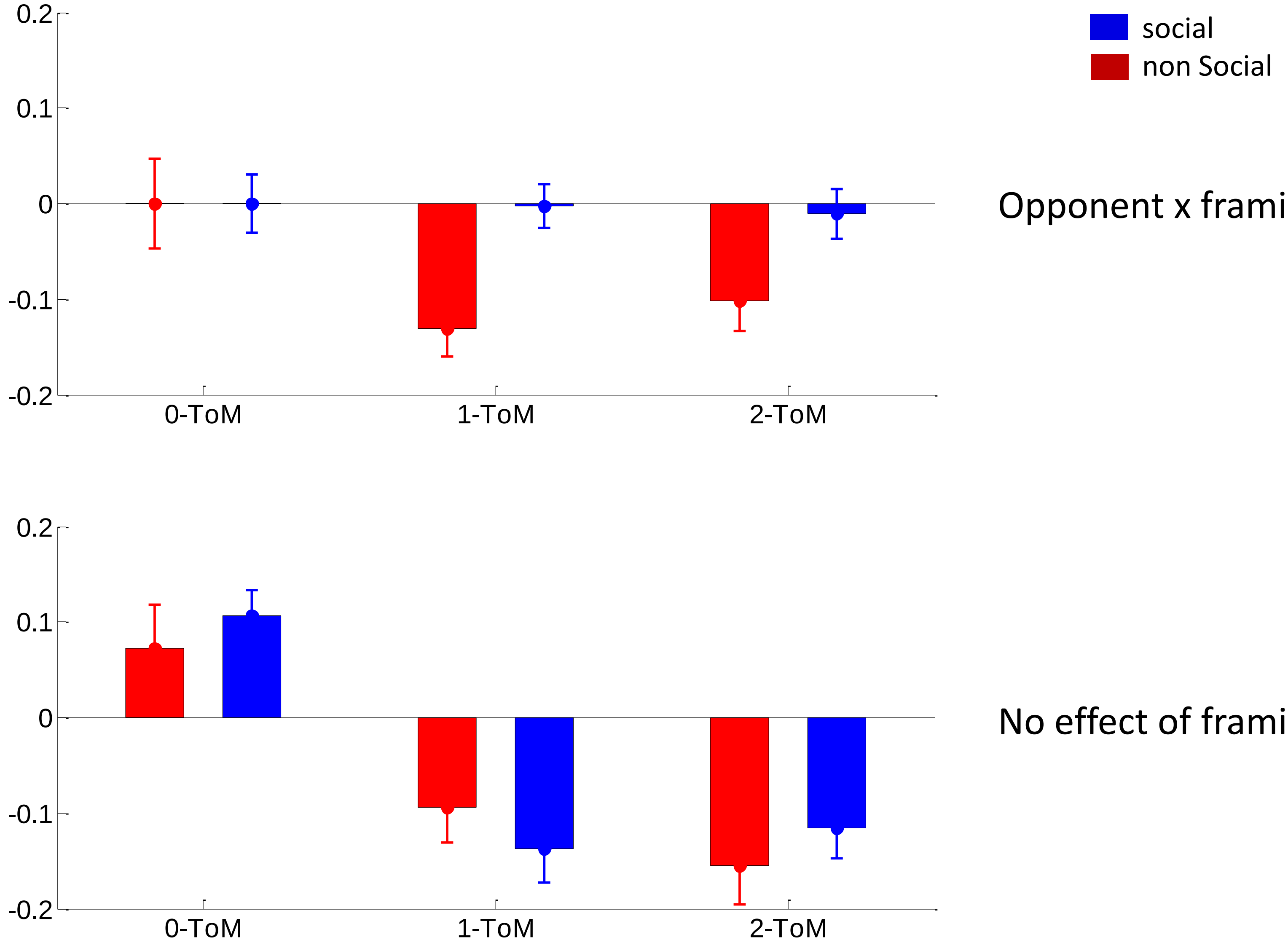
Group	ASD	NT
Age	25,5 (5,7)	27,9 (8,6)
IQ	104(17)	106 (14)
Social anhedonia	14,8 (8,4)	9.7 (4,2)

deceptive framing manipulation: sanity check



Devaine et al. (2015)

Behavioural performances (relative to RB)



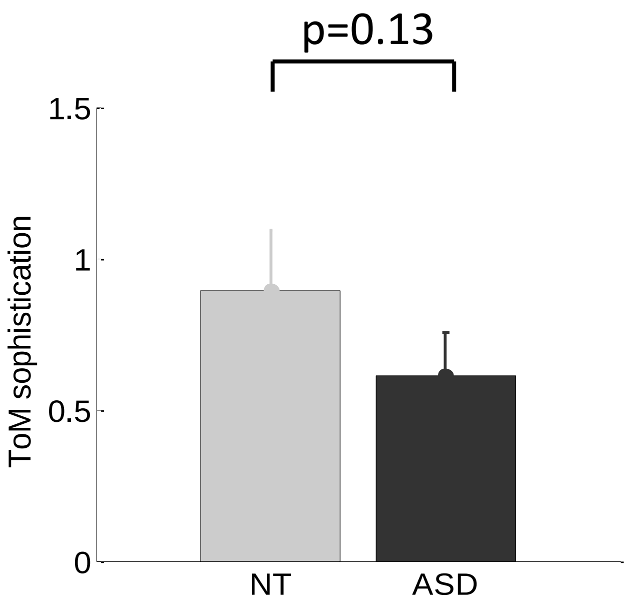
NT — Opponent x framing

ASD — No effect of framing

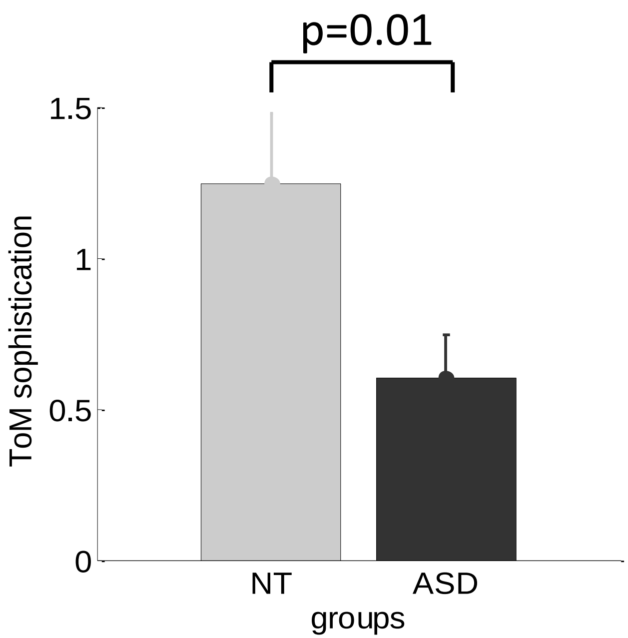
Model-based analysis: ToM sophistication

ToM sophistication: $$E\left[k\middle|y\right]=\sum_{k=0}^{3}k\ p\left(k\middle|y\right)$$

NS framing

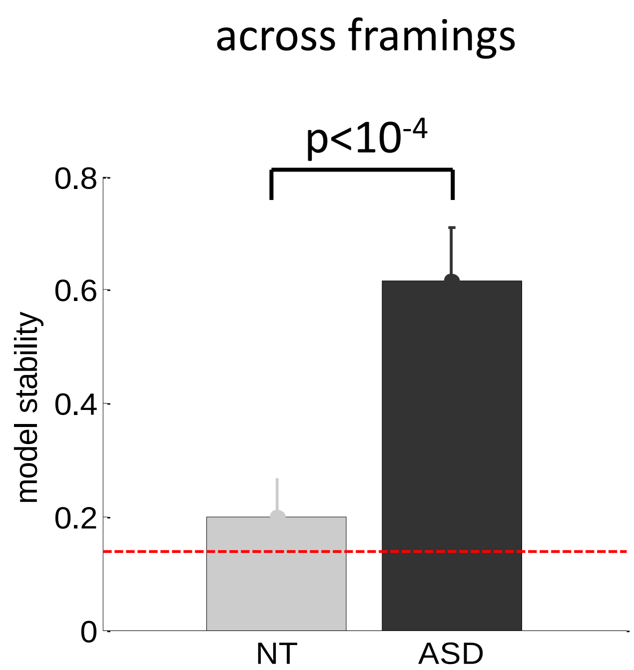


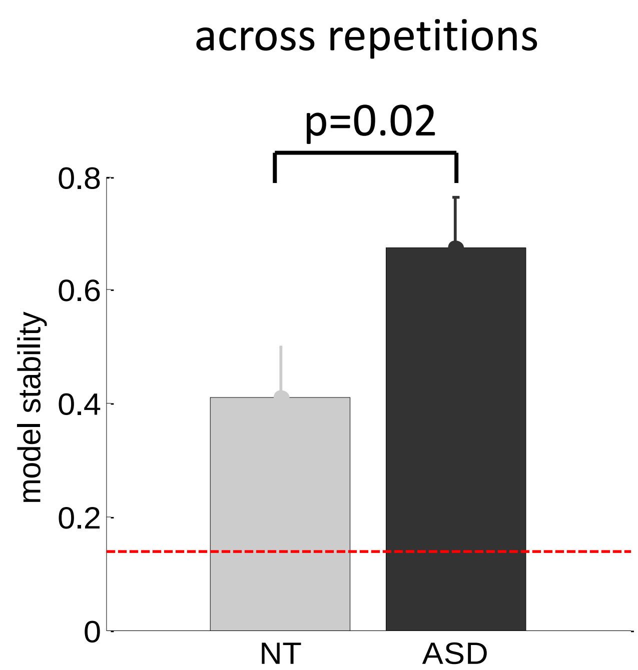
Soc framing



Model-based analysis: learning style rigidity

Model stability: $$P\left(m_1 = m_2 \middle| y\right) = \sum_m P\left(m_1 = m \middle| y_1\right) P\left(m_2 = m \middle| y_2\right)$$



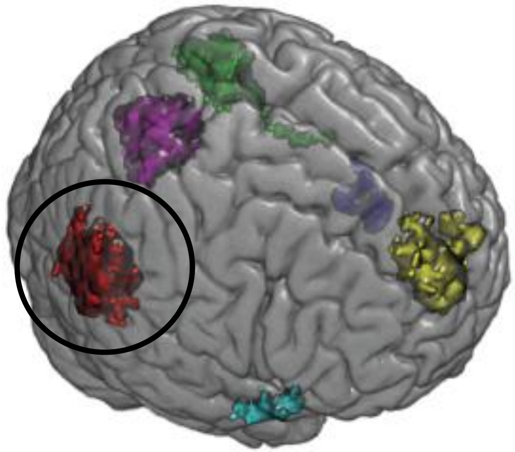


Overview of the talk

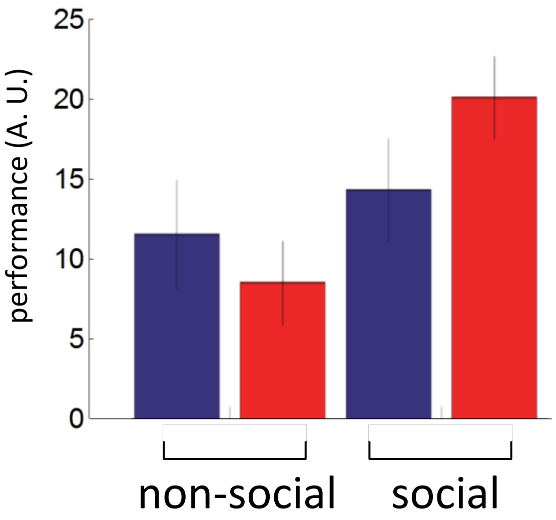
- ✓ Does ToM make a difference when we learn?
- ✓ Limited ToM sophistication: did evolution fool us?
- ✓ Playing *hide-and-seek* with non-human primates
- ✓ What about people with autism spectrum disorder?
- ✓ A short lesson from fMRI

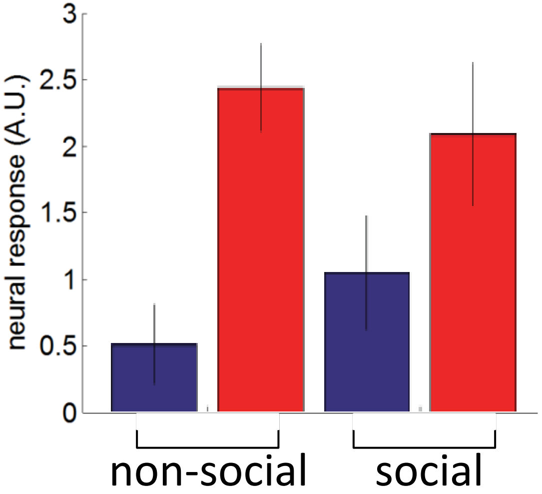
(embarrassingly preliminary) fMRI results

- 2 framings (social VS non-social)
- 2 types of complexity (**sequence** VS **1-ToM**)



temporo-parietal junction





Hill et al. (not even in prep.)

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?
- ✓ A short lesson from fMRI

Summary

- 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:
  - ToM sophistication : ASD < NT
  - Learning style rigidity: ASD > NT
- fMRI:
  - rTPJ: automatically detects intentional behaviour?

References and acknowledgements

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

I also would like to thank:
- Dr. **G. Hollard**, economist (Polytechnique, Palaiseau, France)
- Dr. **S. Masi**, ethologist (Natural History Museum, Paris, France)
- Dr. **B. Forgeot d'Arc**, psychiatrist (Hopital-des-Rivières, Montréal, Canada)
- Dr. **C. Ruff**, neuropsychologist (UZH, zurich, Switzerland)

M. Devaine , A. San-Galli, C. Trapanese, G. Bardino, C. Hano , M. Saint Jalme, S. Bouret, S. Masi, J. Daunizeau (2017), *A computational assay of Theory of Mind sophistication across primate species* . PloS Comp. Biol., in rev.

D. Klindt, M. Devaine, J. Daunizeau (2016), *Does the way we read others' mind change over the lifespan? Insights from a massive web poll of cognitive skills from childhood to late adulthood. Cortex*, in press.

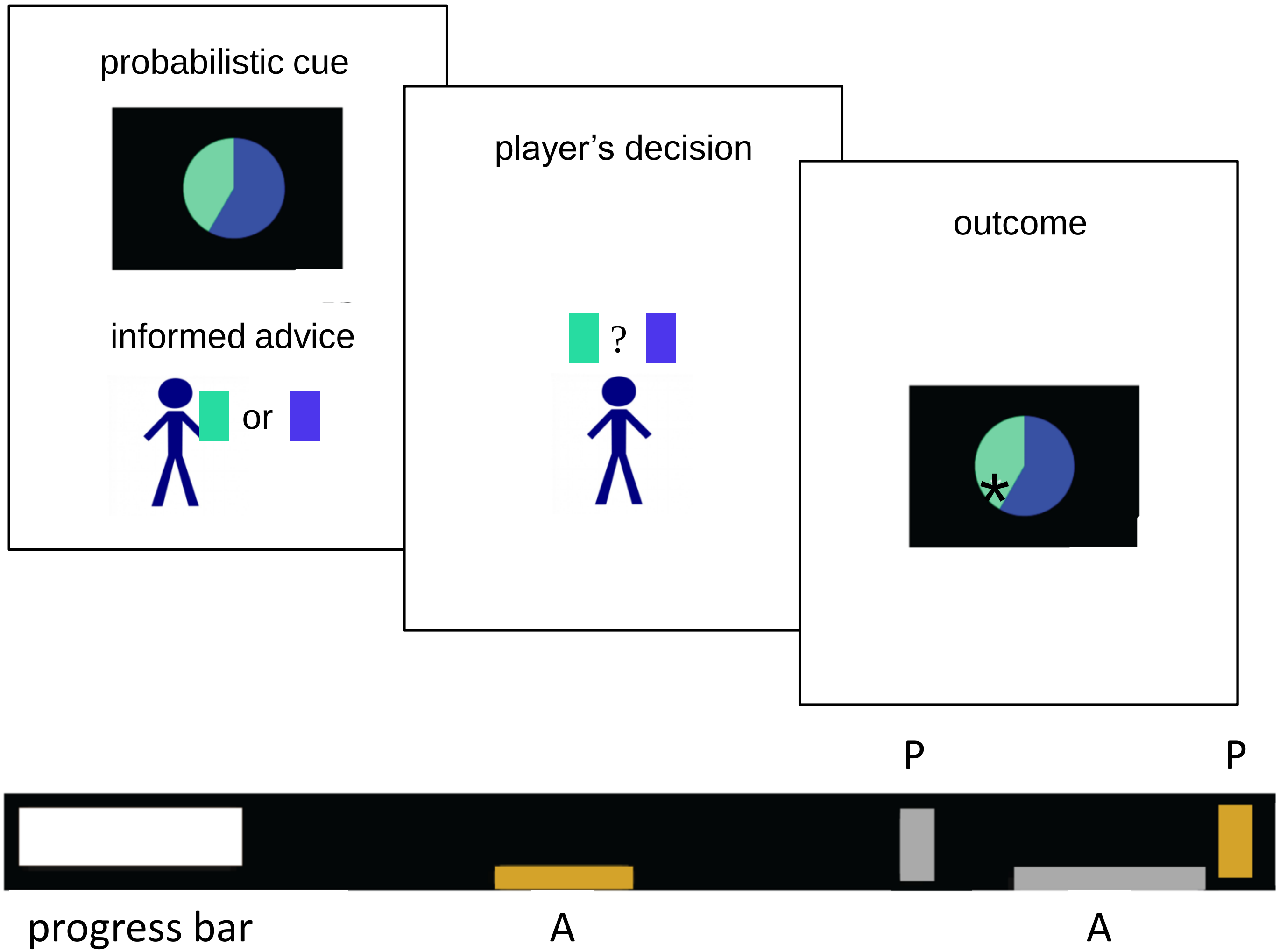
M. Devaine, G. Hollard, J. Daunizeau (2014a), *The social Bayesian brain: does mentalizing matter when we learn?* PLoS Comp. Biol. 10(12): e1003992.

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

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

M. Devaine, A. Duquette, B. Forgeot d'Arc, J. Daunizeau (2015), *Shortcutting reciprocity: using similar, low-recursion learning styles to predict humans and machines behaviour underpins strengths and weaknesses in ASD. A computational psychiatry study.* Int. Meet. Autism Res.

Dealing with uncertain motives: advice taking task



Gold target = 20 CHF
Silver target = 10 CHF

[Diaconescu et al., 2014]

Dealing with uncertain motives: results (N=16)