Özdem et al (2016). Believing androids - fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. *Social Neuroscience*, 7, 1–12.

Ruud Hortensius

Social Robotics Journal Club

19.07.17

SOCIAL ROBOTS

SoBA Lab

SOCIAL BRAIN IN ACTION

PRIFYSGOL BANGOR UNIVERSITY

# Previous literature

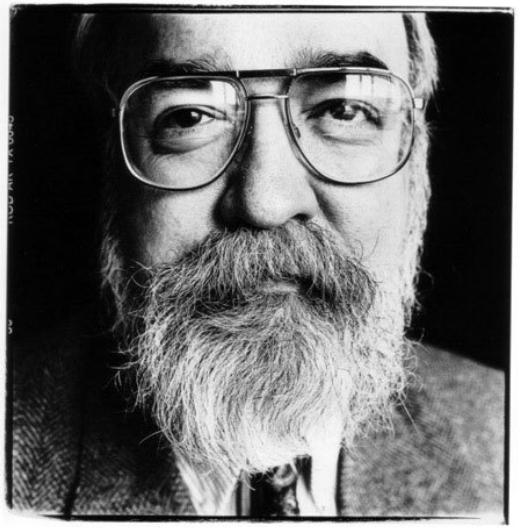
**From metalizing to 'Intentional stance'**

make inferences about the others' internal states (i.e., intentions, beliefs) in order to explain, understand, and predict their behaviour

**Intentional stance:** treating the agent as a rational being with beliefs, desires, and action goals, thereby inducing mentalizing when predicting the agent's behavior.

**Design stance:** explain based on design functionality

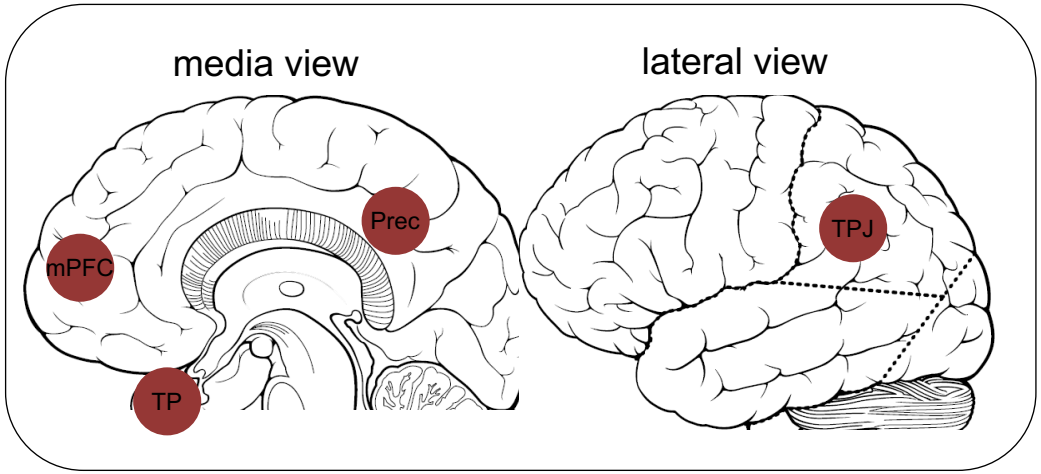
**Physical stance:** explain based laws of physics or chemistry



Dennett, 2003

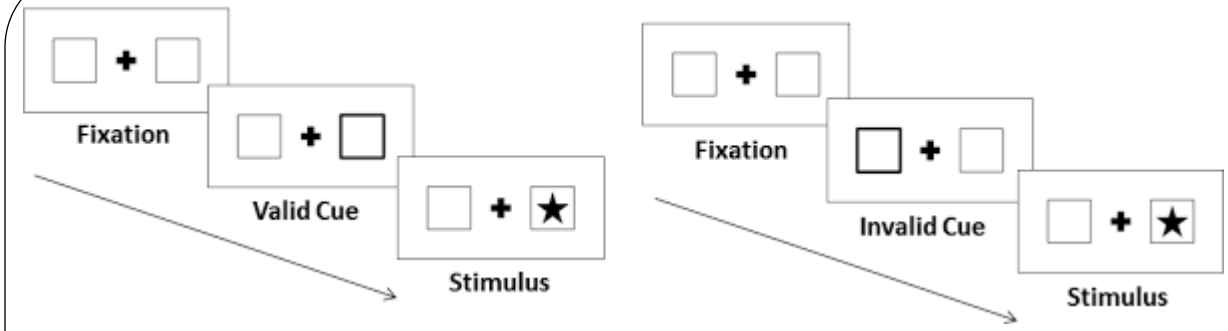
# Previous literature

## Theory-of-mind network



# Previous literature

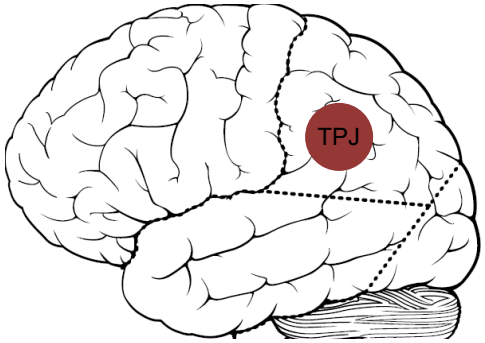
## Attention reorienting




based on Posner, 1980

gaze cueing

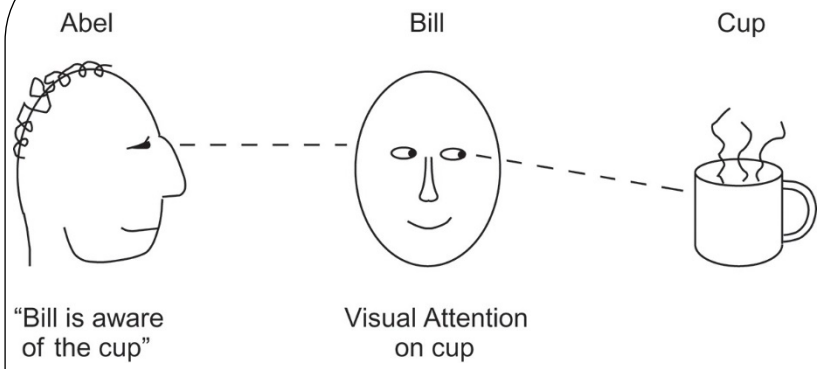
## Neural locus and previous behavioural results



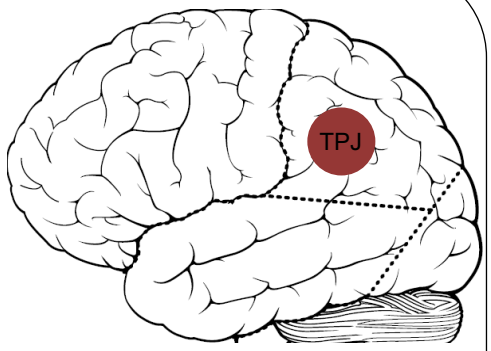

Reorienting
Human > Robots, but based on belief not appearance
Wykowska et al 2014; Wiese et al 2012;

# Aim of the study

## Interaction between mentalizing and attention reorienting

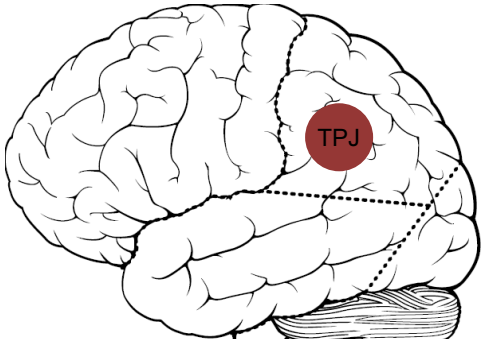



Graziano & Kastner, 2011




pTPJ: mentalizing
aTPJ: attention reorientation
(but maybe also mentalizing)

## Aim of the study




Test aTPJ: ToM x Attention
Behavioural confimation of Wiese et al 2012; Wykowska et al 2014

# Aim of the study

## Expectations

1 “participants would be more likely to orient attention to gazed-at locations and expect to see the target at the gazed-at location when they believe the gaze results from operations of the mind, rather than being algorithmically determined.” → behavioral

2 “Extending on these findings, we expect that the human-controlled condition triggers stronger mind attributions (involving the posterior or anterior TPJ; see Krall et al., 2015) as well as stronger attention to the observed gaze (involving the anterior TPJ) than in the preprogrammed condition.” → main effect of instruction

3 “More importantly, we expect to find interaction effects between cue validity and mind attribution, reflecting larger effects of attentional reorientation after invalid cues (involving the anterior TPJ) when the observed eye gaze is controlled by a human rather than being preprogrammed.” → belief x validity interaction

# Methods

**Procedure adapted from Wykowska et al. 2014**

Gaze cueing task (c.f. Posner)

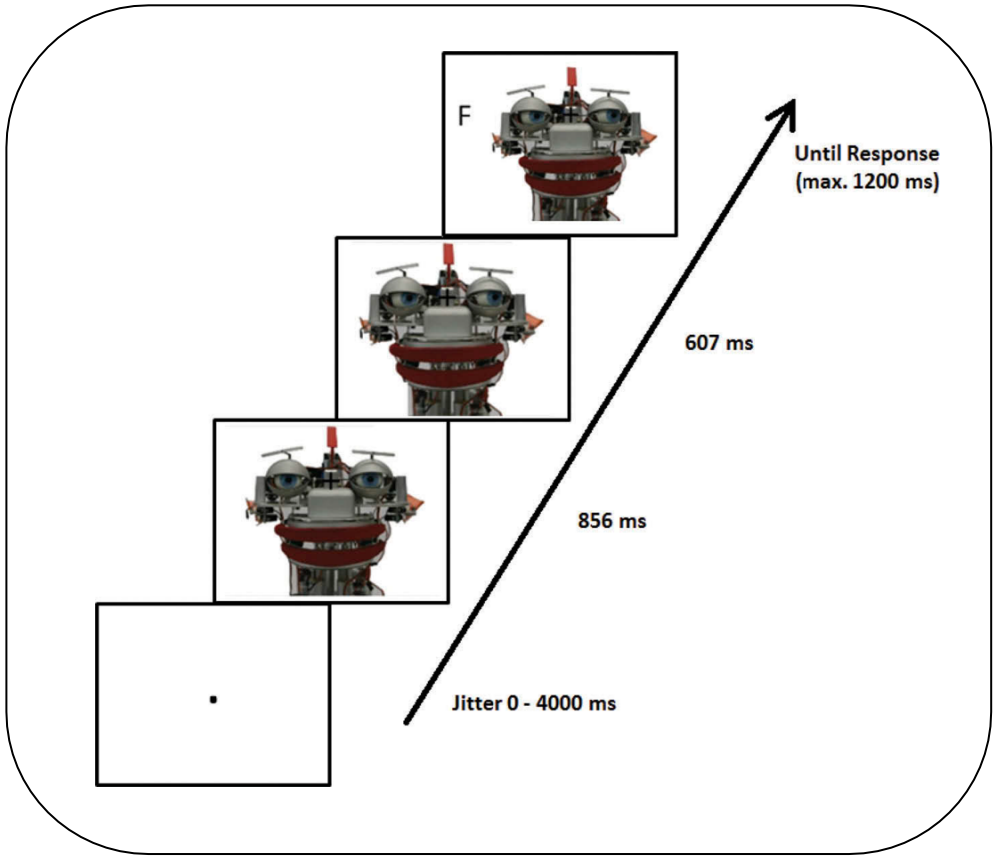
Instruction: respond as quickly and accurately as possible

Valid/non-valid: 50/50

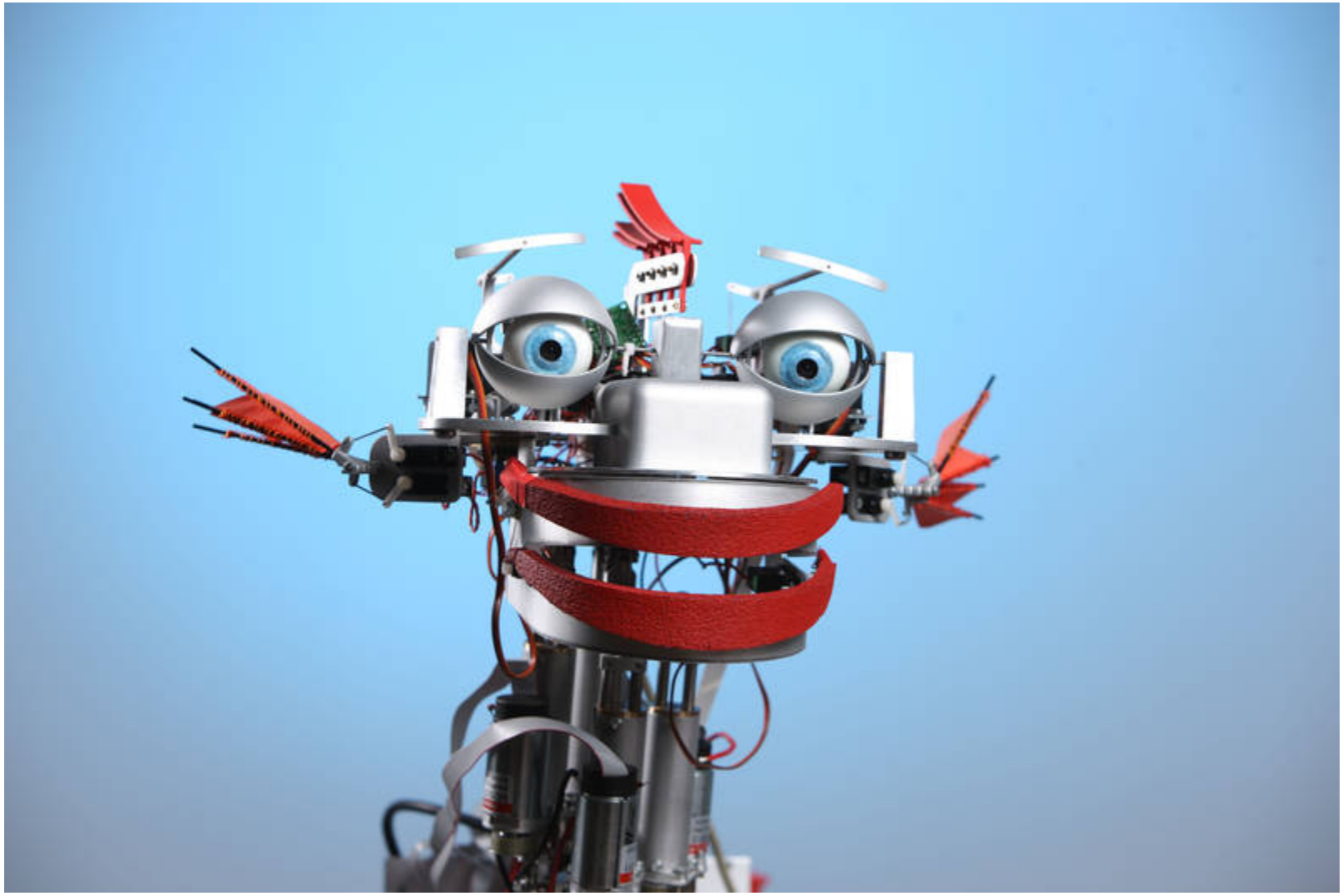
Crucial: effect of instruction on intention attribution only

Belief manipulation:

Human-controlled vs. preprogrammed



# Meet Eddie



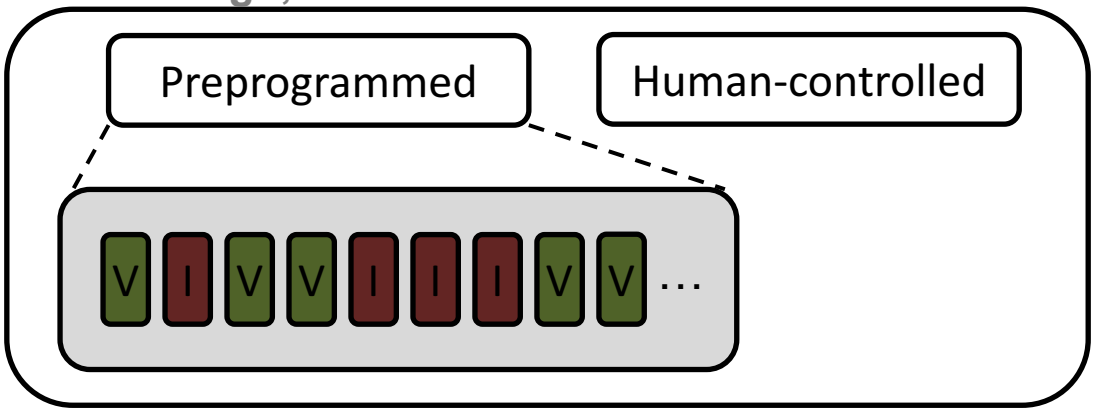
# Procedure

**Belief manipulation**

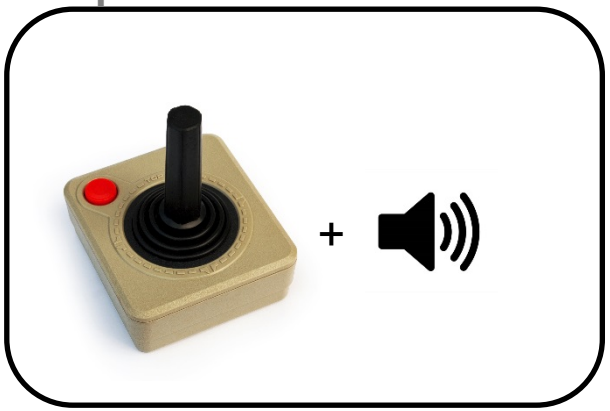
Human-controlled: eye movements of robot controlled by human with joystick

Preprogrammed: eye movements programmed prior to the experiment

**Mixed design, event-related w/ blocks**




**Manipulation**



# Manipulation check

## Funneled questionnaire

1. Was the control of the [preprogrammed] eye movements [by the experimenter] successful?
   If yes, how did you establish that? If no, why did it not work?

2. How good was the [programming of] [control of the experimenter on] the eye movements?
   (1 = not vivid at all to 7 = very vivid)

3. What were the differences between the conditions?
   Please note them all.

4. Did you believe that the [eye movements were preprogrammed] [experimenter had influence on the eye movements of the robot]? If yes, how did you establish that? If no, why did it not work?

5. How large was the influence of the [computer program] [experimenter] on the eye movements?
   (1 = none at all to 7 = complete influence)

6. Did you have any suspicion that the [computer program] [human experimenter] did not really control the eye movements?
   If yes, how did you establish that? If no, why not?

7. How strong was your suspicion?
   (1 = none at all to 7 = very much)

# Data analyse

Final *n* = 21; movement (*n* = 2), suspicion (*n* = 4)

**Behaviour**

2 x 2 ANOVA for $RT_{\text{only correct trials, <2SD}}$ and ACC

**fMRI (whole-brain and ROI analysis)**

Predictors:

4 predictors of interest (valid vs. non-valid; human vs. preprogrammed)

6 motion predictors

**Contrasts**

Main effect of instruction ($H_v + H_i > P_v + P_i$)

Main effect of validity ($H_i + P_i > H_v + P_v$)

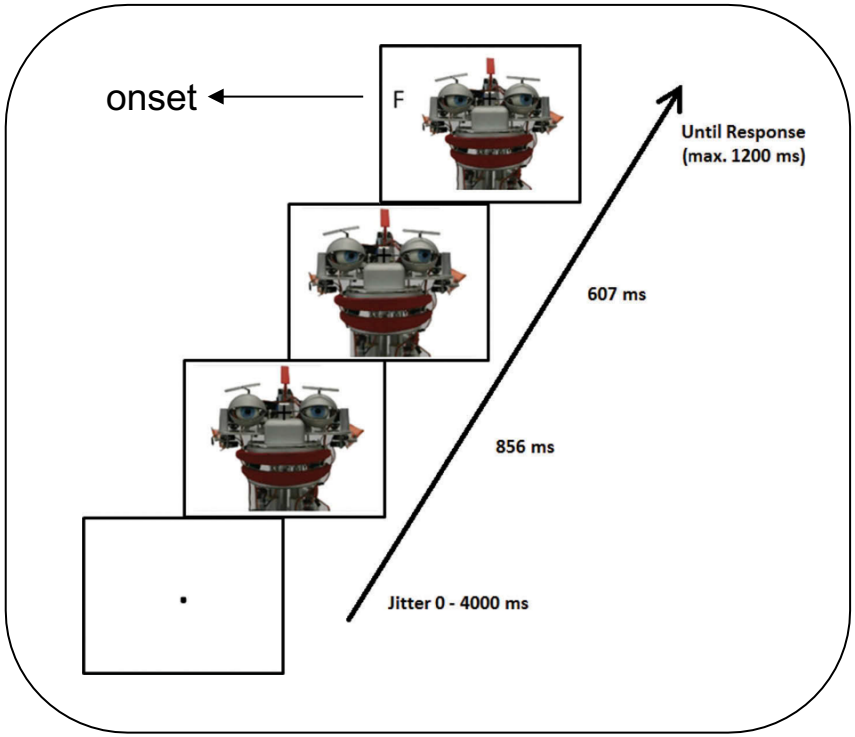
Interaction ($H_i + H_v > P_i > P_v$) → based on RT

(3 -2 1 -2)

**ROIs**

Based on meta-analysis by Bzdok et al (2013):

15 mm sphere (TPJ)



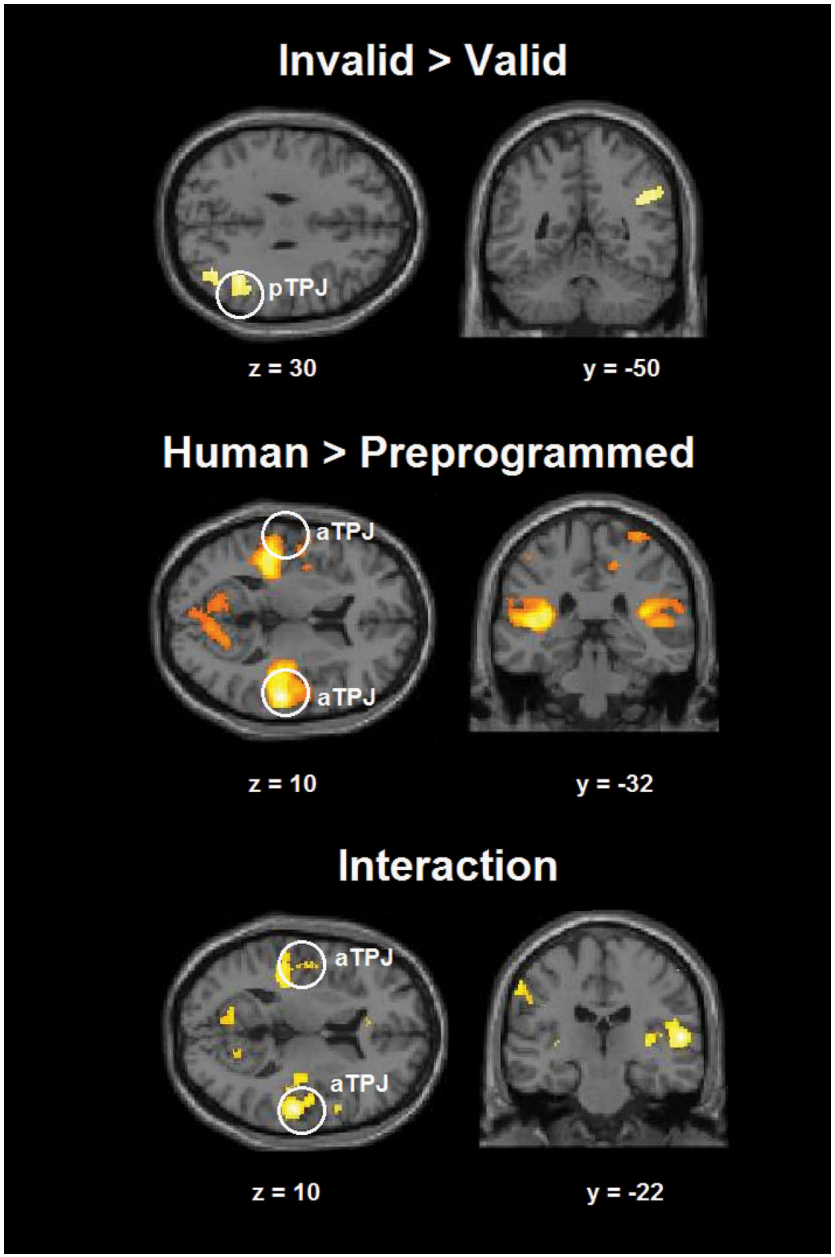
# Behavioural results

| | Human | | Preprogrammed | |
|---|---|---|---|---|
| | Valid | Invalid | Valid | Invalid |
| Response time (ms) | | | | |
| Mean | 504 | 512 | 505 | 510 |
| SD | 24 | 28 | 19 | 24 |
| Accuracy (%) | | | | |
| Mean | 93 | 93 | 93 | 92 |
| SD | 3 | 3 | 3 | 4 |

ANOVAs

| | $F(1,26)$ | $p$ | $np^2$ |
|---|---|---|---|
| Response time | | | |
| Instruction | 0.10 | 0.760 | 0.004 |
| Cue validity | 19.28 | 0.000 | 0.426 |
| Interaction | 1.08 | 0.308 | 0.040 |
| Accuracy | | | |
| Instruction | 1.22 | 0.279 | 0.045 |
| Cue validity | 0.52 | 0.477 | 0.020 |
| Interaction | 0.21 | 0.645 | 0.008 |

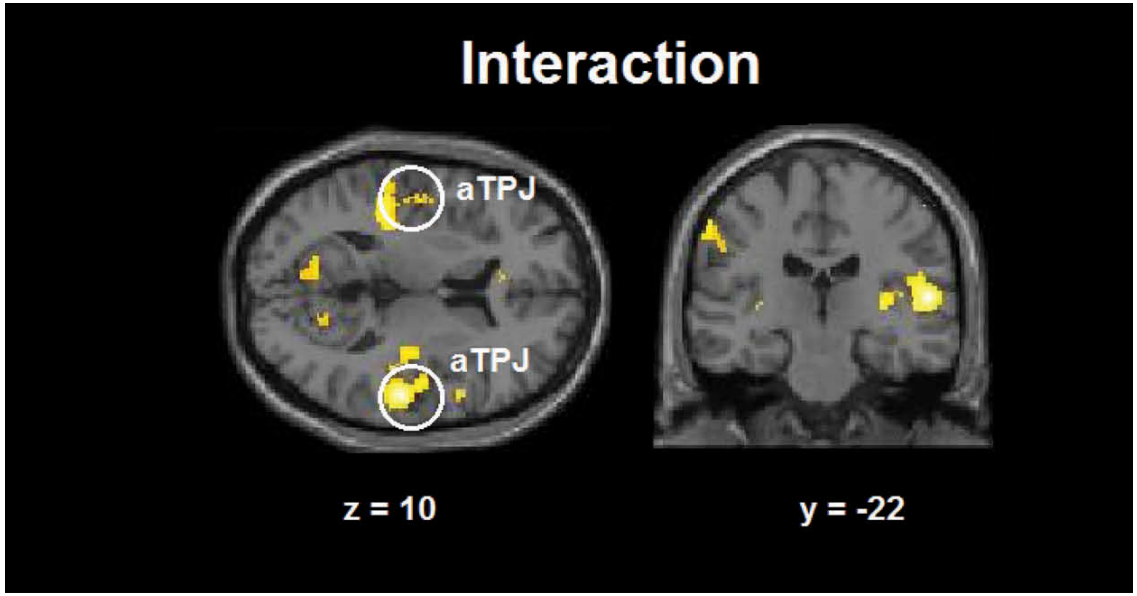
# Whole-brain analysis




*Whole-brain analysis*

| | |
|---|---|
| Invalid > valid | |
| | Right TPJ |
| | Right precuneus |
| Human > preprogrammed | |
| | Right TPJ |
| | Left TPJ |
| | Cuneus |
| | Right superior parietal lobule |
| | Right postcentral gyrus |
| Interaction: human (invalid > valid) > preprogrammed (invalid > valid) | |
| | Right TPJ |
| | Right lentiform nucleus |
| | Right superior temporal gyrus |
| | Left TPJ |
| | Left insula |
| | Left superior temporal gyrus |
| | Left claustrum |

# ROI analysis




*Regions of interest*

Invalid > valid

Right posterior TPJ mentalizing

Human > preprogrammed

Right anterior TPJ reorienting

Left anterior TPJ reorienting

Interaction: human (invalid > valid) > preprogrammed (invalid > valid)

Right anterior TPJ reorienting

Left anterior TPJ reorienting

# Bonus results

instruction main contrast (human > preprogrammed) as well as in the interaction between cue validity and Instruction. We found negative correlations across participants for the bilateral anterior TPJ in the human conditions, which reached significance when pooled together for the right TPJ ($r = -.28$, $p < .05$ one-sided) and the left TPJ ($r = -.41$, $p < .01$ one-sided). Thus, the stronger the activation in the anterior TPJ (main effect of instruction or interaction between cue validity and instruction), the faster participants responded in the human conditions. There were no significant correlations for the preprogrammed conditions.

But one-sided and not corrected for multiple comparisons (4)

effect and mind attribution was not significant. Note, however, that there were sizeable correlations between this interaction contrast and response times, indicating that participants who showed stronger contrast activations in the anterior TPJ responded faster in the human-controlled conditions. The lack of a statistically significant

# Discussion

## FIndings

1 “participants would be more likely to orient attention to gazed-at locations and expect to see the target at the gazed-at location when they believe the gaze results from operations of the mind, rather than being algorithmically determined.” → behavioral

2 “Extending on these findings, we expect that the human-controlled condition triggers stronger mind attributions (involving the posterior or anterior TPJ; see Krall et al., 2015) as well as stronger attention to the observed gaze (involving the anterior TPJ) than in the preprogrammed condition.” → main effect of instruction

3 “More importantly, we expect to find interaction effects between cue validity and mind attribution, reflecting larger effects of attentional reorientation after invalid cues (involving the anterior TPJ) when the observed eye gaze is controlled by a human rather than being preprogrammed.” → instruction x validity interaction

# Discussion

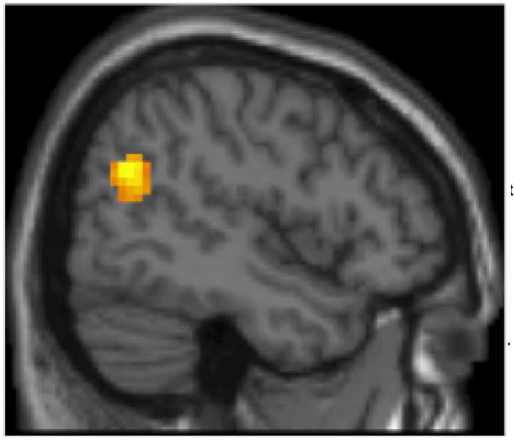
**aTPJ**

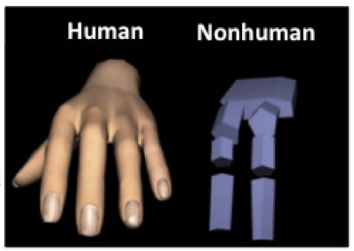
mentalizing + reorienting

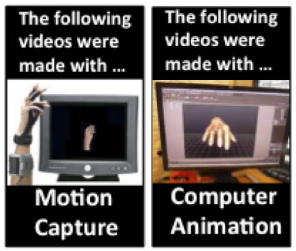
High-level: mentalizing

↕

Low-level: attention




NB. rTPJ → dMPFC connectivity

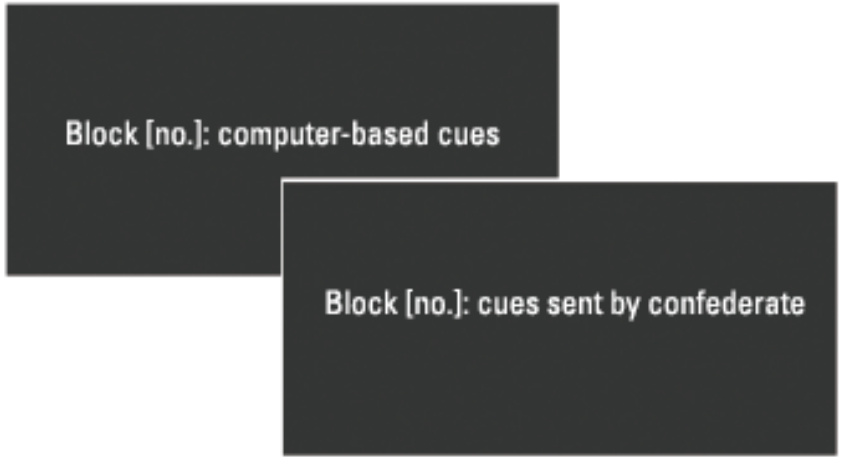
Klapper, Ramsey et al 2014

Engagement of TPJ when cues of animacy are present

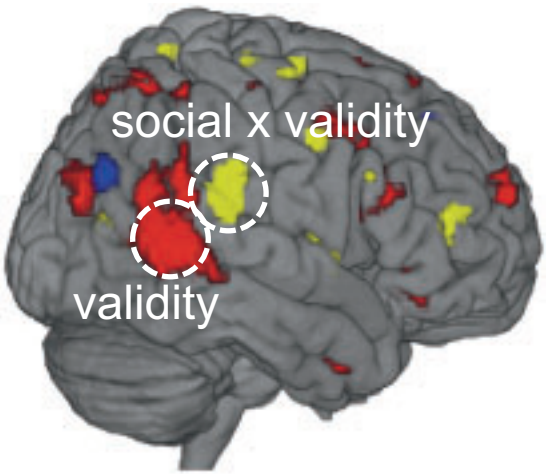
*But what about pTPJ?*

*Impact of interaction?*

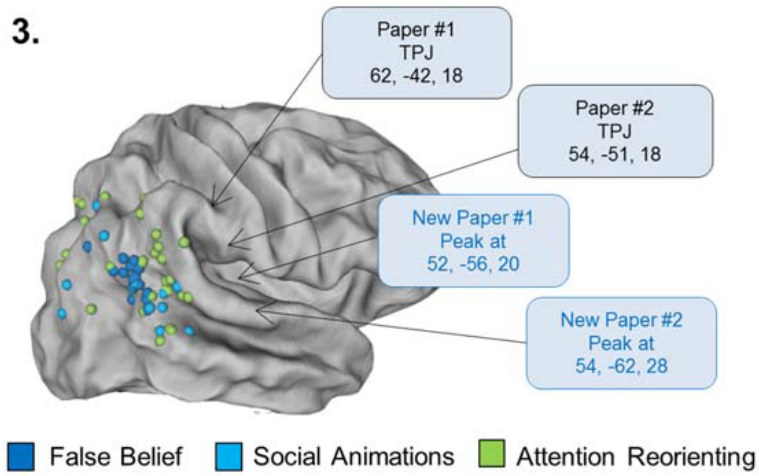
# Reflection

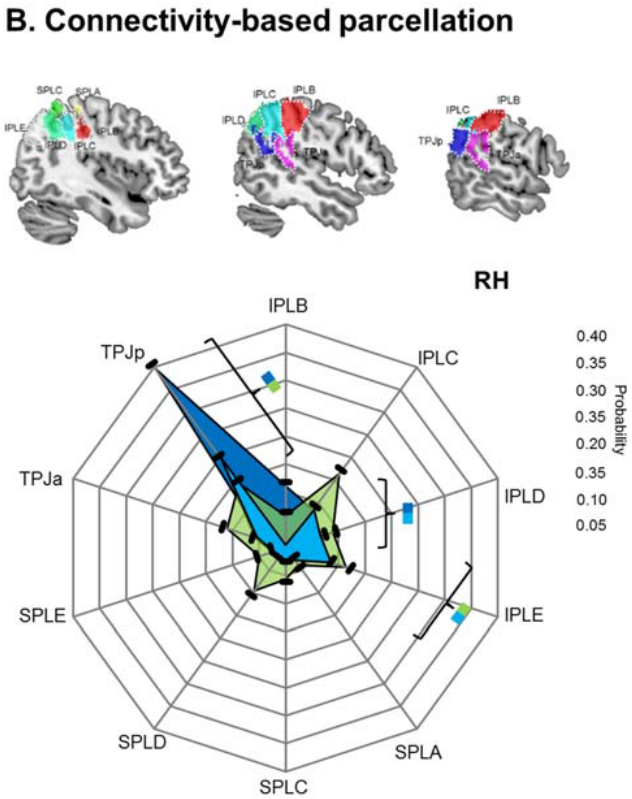



Schuwerk et al 2017



# Reflection



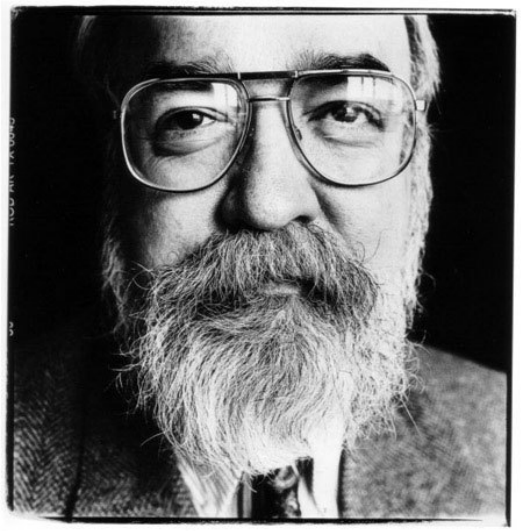

Schurz et al 2017

# Reflection

Intentional stance?

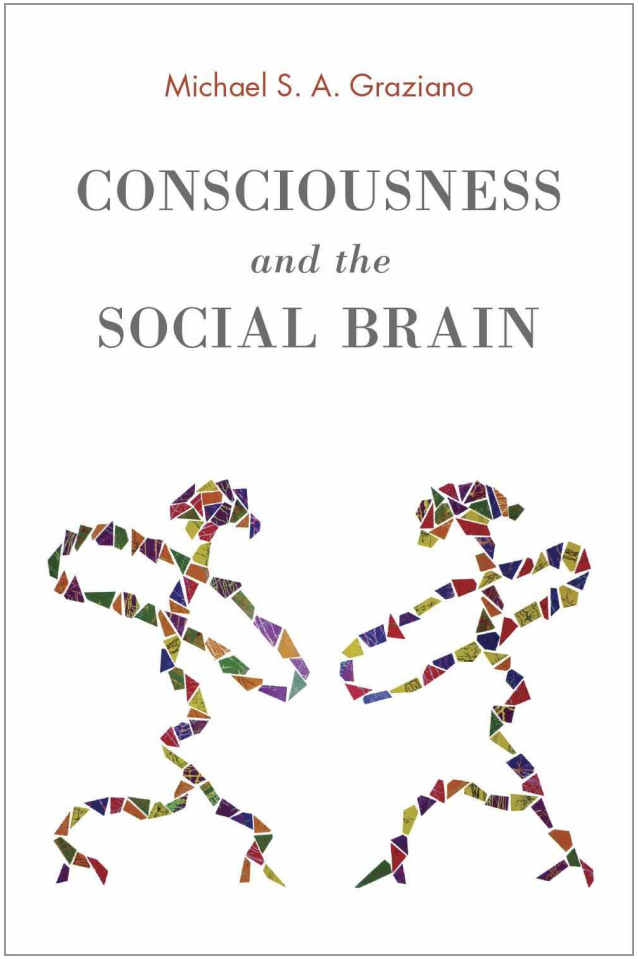
or

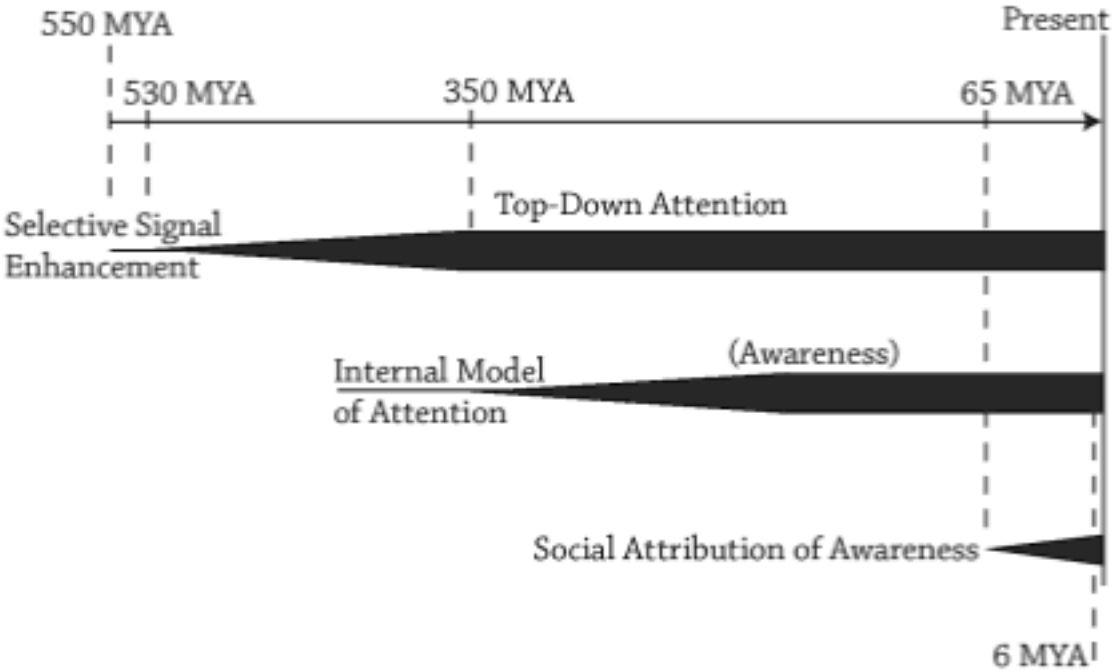
design stance?



Dennett, 2003

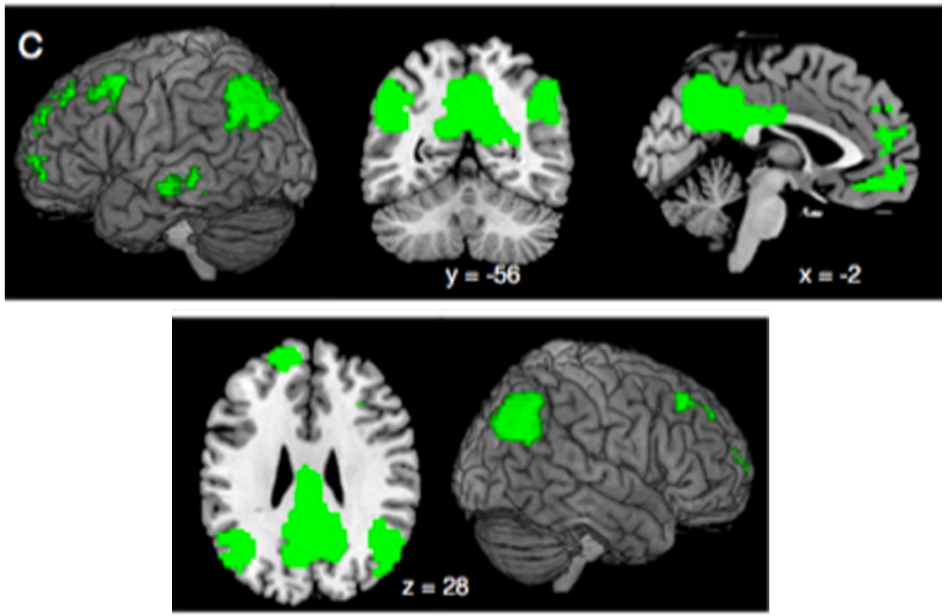
# Going beyond the scope



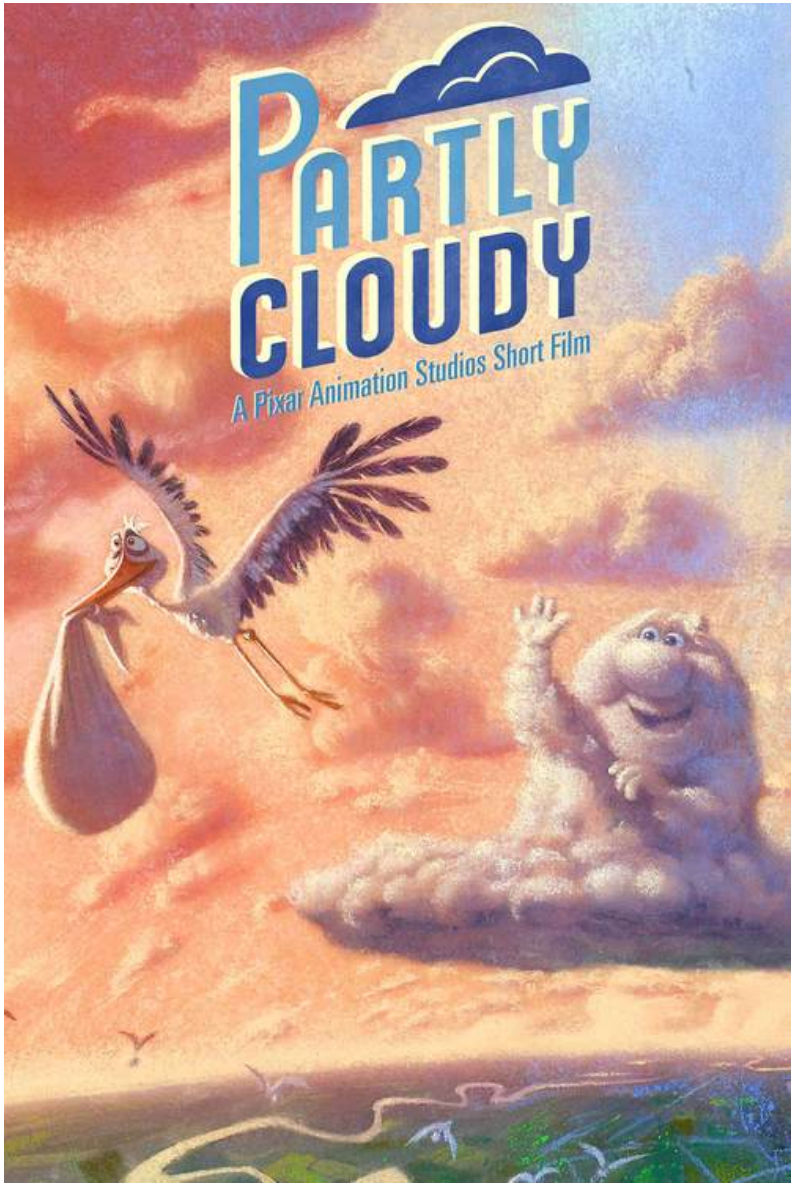



# Reflection

- no functional ROIs




Jacoby et al 2016



# Discussion

Accordingly, our finding suggests that participants paid increased attention to the face and gaze of the robot under the belief that the robot's eyes were human-controlled because changes in gaze direction coming from a human partner would constitute a more important visual stimulus than those generated by a machine. This interpretation is in line with what participants reported in the postscanning suspicion questionnaire.

But...

| | $F(1,26)$ | $p$ | $np^2$ |
|---|---|---|---|
| Response time | | | |
| Instruction | 0.10 | 0.760 | 0.004 |
| Cue validity | 19.28 | 0.000 | 0.426 |
| Interaction | 1.08 | 0.308 | 0.040 |
| Accuracy | | | |
| Instruction | 1.22 | 0.279 | 0.045 |
| Cue validity | 0.52 | 0.477 | 0.020 |
| Interaction | 0.21 | 0.645 | 0.008 |

# Discussion

- no behavioural effect (low number of trials/scanning environment and parameters/online paradigm)

- 384 trials in total, 96 per condition (e.g. human_valid)

- *Manipulation too subtle?*

# Reflection

- voxel size: 3.5 isotropic, gap (distance factor)
- no post scanning questionnaire results shown
- is the sound a cue for humanness? <— authors argue against this
- connectivity?
- ascribe function to region without behavioural confirmation
- non-social control?

# Conclusion

To summarize, the goal of the present study was to investigate the neural correlates underlying the effect of mind attribution on attentional reorienting to gaze direction. We found that higher-order social attribution processes and lower-level attention mechanisms intersect at the anterior TPJ responsible for attentional reorientation and belief inference (Krall et al., 2015), which exhibited particularly strong activation for invalid cues under the belief that eye movements were human-controlled. Based on our findings, gaze cues believed to be produced by a human rather than a machine seem to attract more attention and most likely reflect the fact that they are socially more relevant and informative than preprogrammed ones. This adds to the growing evidence that mind attribution and attentional orienting interact in a common, underlying attentional control process (Mitchell, 2008; Özdem et al., 2016; Scholz et al., 2009), which ensures that more attentional resources are devoted to interactions with agents who are believed to have a mind, as compared to machine-like agents.