



Universidad de Valladolid

Fundación
BBVA



PROYECTO
KEAMA

UNDERSTANDING INTENTIONALITY ATTRIBUTION IN AUTISM: EXPLORING THE KNOBE EFFECT ACROSS DOMAINS

2ND KEAMA WORKSHOP: MORAL COGNITION AND COMMUNICATION IN AUTISM

UNIVERSITY OF VALLADOLID

8TH MARCH, 2024

AGUSTÍN VICENTE BENITO
IKERBASQUE & UPV/EHU
agustin.vicente@ehu.eus

JOSÉ V. HERNÁNDEZ-CONDE
UNIVERSITY OF VALLADOLID
jhercon@uva.es

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

CONCLUSIONS

AGENDA

I. INTRODUCTION

- The Knobe effect
- Morals and intentionality attribution in autism

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

- Attribution of intentionality
- Attribution of praise and blame

IV. DISCUSSION OF RESULTS

- Intentionality attribution
- Moral evaluation
- General comparison of autistic vs neurotypical populations
- Back to the Knobe effect

CONCLUSIONS

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

CONCLUSIONS

AGENDA

AGENDA

I. INTRODUCTION

- THE KNOBE EFFECT
- MORALS AND INTENTIONALITY ATTRIBUTION IN AUTISM

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

CONCLUSIONS

WHAT IS THE KNOBE EFFECT?

The Knobe effect (or side-effect effect) is the phenomenon by which people tend to judge

- that a negative side effect has been intentionally provoked
- that a positive side effect has not been intentionally provoked

This means that ...

- ... people attribute intentionality to actions depending on whether they perceive the outcome of the action as positive or negative
- ... attributing more intentionality to actions with negative effects (e.g., to harm) than to actions with positive moral valence (e.g., to help)

Consequently ...

- ... people's perceptions of the moral character of an action have an influence on their attributions of intentionality

DISCOVERY, INFLUENCE AND IMPORTANCE

The Knobe effect is identified by Joshua Knobe in 2003

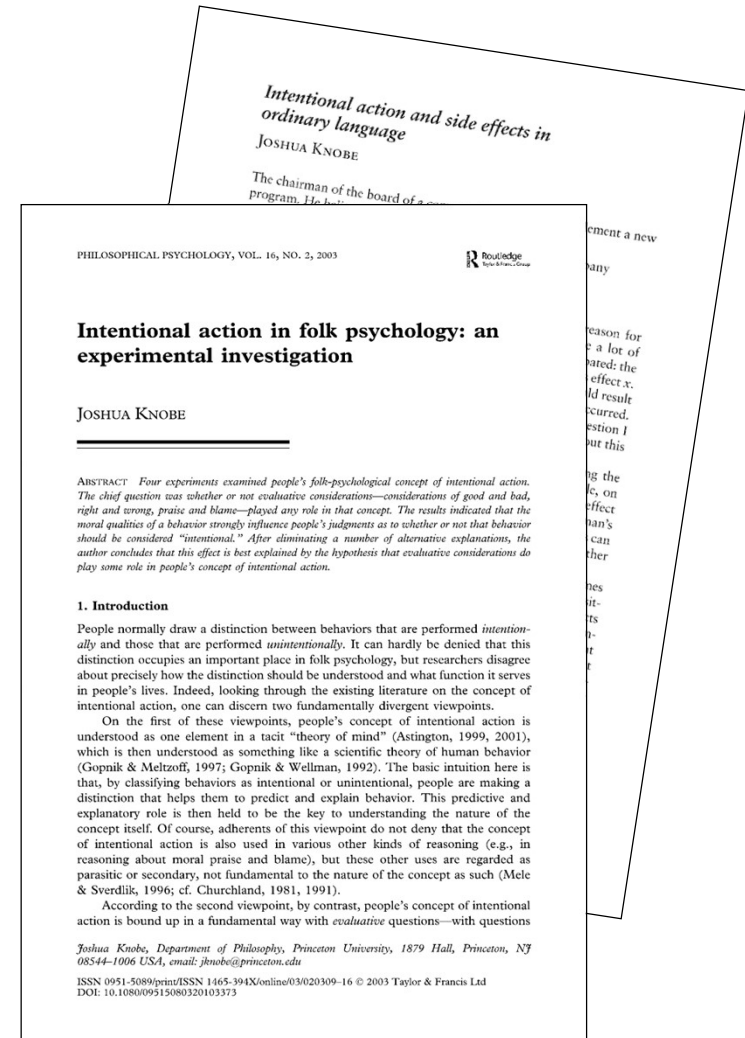
- ... for the attribution of intentionality in cases with moral valence⁽¹⁾

Since then, it has had a great influence

- 985 articles in PhilPapers

The Knobe effect has been found in many different areas

- Attribution of intentionality
- Attribution of belief
- Attribution of knowledge
- Attribution of agency / group responsibility
- Attribution of externalities



(1) Earlier, Harman (1976) had suggested that a person's intuitions regarding the intentionality of a certain side effect could be influenced by that person's attitude toward that specific side effect

SITUATIONS AND CONTEXTS WHERE IT HAS BEEN IDENTIFIED

The Knobe effect in the moral domain has been identified

- ... in other cultural and geographic environments (Knobe & Burra 2006; Kaspar *et al.* 2016)
- ... in young children (Leslie *et al.* 2006; Pellizzoni *et al.* 2009)
- ... in people with brain lesions in the prefrontal cortex (Young *et al.* 2006)
- ... in subjects with autism spectrum disorder (Zalla & Leboyer 2011)
- ... and by XPhil replicability project (Cova *et al.* 2021)

It has also been observed and/or applied in areas other than morality

- *Epistemological*: attribution of belief (Beebe 2013) and knowledge (Beebe & Buckwalter 2010)
- *Economic / business* (Wible 2008; Utikal & Fischbacher 2014)
- *Legal* (Cohen-Eliya & Porat 2015)

DOMINANT INTERPRETATION OF THE KNOBE EFFECT

(1) **Moral valence** (Knobe 2003a, 2003b)

The moral valence (i.e., character + or – of the side effect) alone is responsible for the Knobe effect

Evidence of this would be that this asymmetry is found ...

- if *having the intention of* is changed by ...
- *pretending, deciding, defending and opposing* (Pettit & Knobe 2009)

Problem

- moral valence does not fully explain the Knobe effect, as ...
- ... there are cases where subjects attribute intentionality to the agent, although the side effect is not negative (Phelan & Sarkissian 2009)

In any case, the interpretation based on moral valence remains the reference explanation

ALTERNATIVE EXPLANATIONS (I)

(2) Explanatory diversity (Nichols & Ulatowski 2007)

Two different concepts of intentional action come into play in Knobe's scenarios

- one based on *desires*
- one based on *beliefs*

(3) Trade-off hypothesis (Machery 2008)

People think that the agent provokes a negative side effect in exchange for a benefit⁽¹⁾

(4) Pragmatist explanations (Adams & Steadman 2004a, 2004b)

When subjects ascribe intentionality to the agent, they blame the agent's decision (by means of a conversational implicature)

(1) This would only be the case in cases with negative side effects, as there are no costs associated with those with positive consequences

ALTERNATIVE EXPLANATIONS (II)

(5) **Biased scenarios** (Nadelhoffer 2004, 2006)

The attribution of intentionality is biased by the way the two scenarios (with + and – effects) are described

- The negative scenario introduces a negative impression of indifference / unconcern on the part of the agent

(6) **Normativist explanations** (Holton 2010; Alfano *et al.* 2012)

The asymmetry is due to the fact that people form stronger beliefs in cases with negative side effects ...

- ... for being cases of violation of rules

(7) **Probabilistic explanations** (Dalbauer & Hergovich 2013)

The asymmetry is due to the different perception of the probability that the action will cause the side effect (for the + and – cases)

AGENDA

AGENDA

I. INTRODUCTION

- THE KNOBE EFFECT
- MORALS AND INTENTIONALITY ATTRIBUTION IN AUTISM

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

CONCLUSIONS

MORAL DEVELOPMENT AND PSYCHOLOGY IN AUTISM (I)

Autistic children have been found to

- ... be less strict about the moral / conventional distinction than neurotypical children (Shulman *et al.*, 2012)
- ... be less elaborate in their judgments (*ibid.*)
- ... have difficulty distinguishing bad outcomes of intentional vs. accidental actions from a moral perspective
- ... tend to judge more in terms of outcomes than of intentions (Moran *et al.*, 2011; Margoni & Surian, 2016)

Since ToM difficulties in autism are thought to persist into adulthood

- ... it is assumed that autistic adults will show a stronger tendency than neurotypicals to judge the morality of an action by its consequences
- ... especially in cases where tracking the intentions of agents may be complicated (Garcia-Molina & Clemente-Estevan, 2019)

MORAL DEVELOPMENT AND PSYCHOLOGY IN AUTISM (II)

Zalla & Leboyer's (2011) studies with autistic adults have provided some support for this idea

- ... their blame/praise judgment was less related to intentionality attribution than in the neurotypical population
- ... suggesting that the autistic people did not evaluate agents' actions on the basis of their intentions

In addition, autistic individuals may exhibit a stronger sense of justice or more consistent morals than neurotypicals (Dempsey *et al.*, 2020)

- Some studies suggest that autistics are more legalistic; while neurotypicals are more prone to accept exceptions to moral norms (Strang *et al.*, 2017)
- ... and many autistic adults are irritated by what they perceive as the moral laxity of neurotypicals (Hu *et al.*, 2021)

Their stronger commitment to moral rules may be related to a stronger commitment to rule-based behavior –or to rigid dichotomous thinking– (Petroliini *et al.*, 2023)

OUR HYPOTHESES ABOUT COMPARISONS BETWEEN AUTISTIC AND NEUROTYPICAL ADULTS

- (1)** There will be differences in the moral evaluation of the agents
 - Larger discrepancy in attributed praise / blame
 - Autistics will be stricter in their judgments than neurotypicals
 - Higher standards in the praise / blame question for the autistic group

- (2)** There will be differences in intentionality attribution to agents
 - Larger discrepancy in intentionality attributions for the autistic group
 - Larger Knobe effect for autistics
 - ... i.e., autistic individuals will be more prone to the Knobe effect

- (3)** Autistic individuals may base on outcomes
 - ... their intentionality attributions
 - ... but not their moral judgments

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

CONCLUSIONS

PRESENTATION OF THE EXPERIMENT

The experiment is based on Knobe's (2003a) work

- ... for the *moral* domain ...
- ... with alternative cases in the *aesthetic* and *alethic* domains

For each domain, 2 critical vignettes were used

- ... 1 with a *positive* and
- ... 1 with a *negative* collateral effect

Participants were asked to indicate

- ... the praise / blame attributed to agent X
 - (on a 7-point Likert scale)
- ... whether they thought that agent X intentionally contributed to the occurred side effect

EXPERIMENTAL DESIGN AND PROCEDURE

Each participant was exposed to 1 critical vignette from each domain

- ... along with 6 filler scenarios (that acted as distractors)

The 9 scenarios were organized into 3 consecutive blocks

- Each block consisted of 3 vignettes
 - 1 critical scenario of moral, aesthetic or alethic type
 - with a positive or negative collateral effect
 - 2 distractor scenarios

The other experimental elements were randomized

- Order of presentation of the 3 blocks
- Internal order of vignettes within each block
- Allocation of the 6 filler vignettes to the blocks

The scenarios were shown to the participants in Spanish or English

CRITICAL VIGNETTES (I)

(1) MORAL domain: Knobe's *help / harm* environment vignette

The vice-president of a company went to the chairman of the board and said, “We are thinking of starting a new program. It will help us increase profits, *and/but* it will also *help/harm* the environment.” The chairman of the board answered, “I don’t care at all about *helping/harming* the environment. I just want to make as much profit as I can. Let’s start the new program.” They started the new program. Sure enough, the environment was *helped/harmed*.

[1] Rate your agreement (+3) or disagreement (−3) with the following statement:

“The chairman did the right thing by starting the new program.”

[2] Did the chairman intentionally *help/harm* the environment?

CRITICAL VIGNETTES (II)

(2) ALETHIC domain: *right / wrong* economic outlook vignette

The campaign manager of a political party went to the candidate and said, “We are thinking of using a new argument in the campaign. It will help us respond to the opposition, but it will also make people get the *right/wrong* idea about the economic outlook.” The candidate answered, “I do not care at all about voters getting the *right/wrong* idea about the economic outlook. I just want to respond to the opposition. Let’s use the new argument.” The campaign staff used the new argument. Sure enough, the voters got the *right/wrong* idea about the economic outlook.

[1] Rate your agreement (+3) or disagreement (−3) with the following statement:

“The candidate did the right thing by using the new argument.”

[2] Did the candidate intentionally contribute to the voters getting the *right/wrong* idea about the economic outlook?

CRITICAL VIGNETTES (III)

(3) AESTHETIC domain: *beautiful / ugly* street furniture vignette

The urban planning councilor of a city went to the mayor and said, “We are thinking of replacing the street furniture with new ones. It will give the city a more modern look, but it will also *improve/worsen* the general aesthetics of the city.” The mayor of the city answered, “I do not care at all about *improving/worsening* the aesthetics of the city. I just want it to look more modern. Let’s install the new street furniture.” The city council replaced the old street furniture with the new ones. Sure enough, the general aesthetics of the city *improved/worsened*.

[1] Rate your agreement (+3) or disagreement (− 3) with the following statement:

“The mayor did the right thing by replacing the old street furniture with the new ones.”

[2] Did the mayor intentionally *improve/worsen* the aesthetics of the city?

SELECTION OF PARTICIPANTS

99 neurotypical participants

- ... including both university students and professors
- ... from 3 public universities in Spain
- ... 65% female, average age = 37
- ... predominantly native Spanish speakers
- ... none of them were paid or received course credit for their participation

99 autistic participants

- ... recruited through the Prolific platform
- ... 49% female, average age = 37
- ... native English speakers
- ... each participant received 3£ as compensation

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

CONCLUSIONS

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

- ATTRIBUTION OF INTENTIONALITY
- ATTRIBUTION OF PRAISE AND BLAME

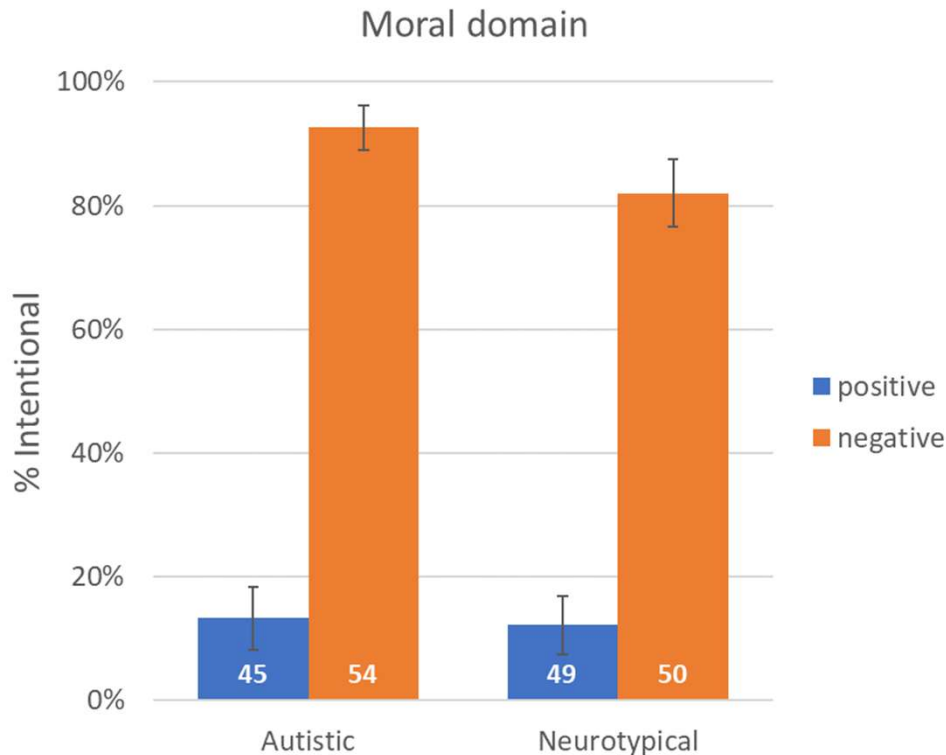
IV. DISCUSSION OF RESULTS

CONCLUSIONS

THE KNOBE EFFECT WAS PRESENT IN THE MORAL DOMAIN FOR BOTH THE AUTISTIC AND NEUROTYPICAL POPULATIONS

As expected, based on the results of previous studies

Attribution of intentionality (in the moral domain)



Differences between the positive (+) and negative (-) side-effect cases

Moral domain: *autistic population*

- Results comparable to Zalla & Leboyer (2011)
 - 92% (-) vs 13% (+)
- Highly significant difference
 - $\chi^2(1, N=99)=59.6, p<0.001$

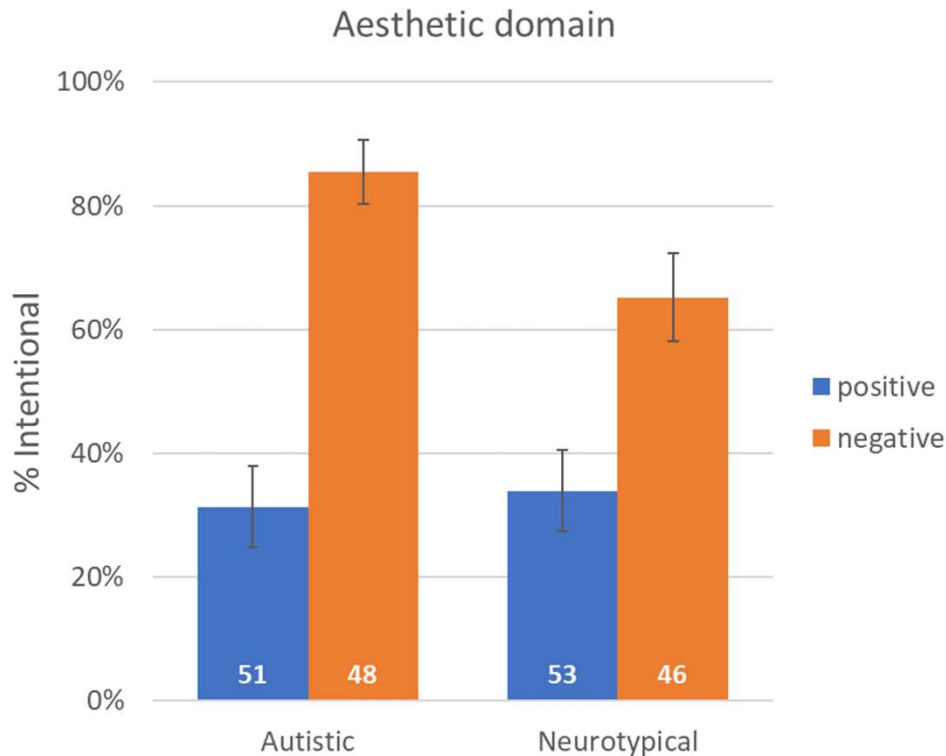
Moral domain: *neurotypical population*

- Results comparable to Knobe's (2003a), and Zalla & Leboyer (2011)
 - 82% (-) vs 12% (+)
- Highly significant difference
 - $\chi^2(1, N=99)=45.5, p<0.001$

IN THE AESTHETIC DOMAIN, INTENTIONALITY ATTRIBUTION IS SMALLER / LARGER IN NEGATIVE / POSITIVE CASES

Nevertheless, the Knobe effect is still present in both populations

Attribution of intentionality (in the aesthetic domain)



Differences between the positive (+) and negative (-) side-effect cases

Aesthetic domain: *autistic population*

- Asymmetry in the attribution of intentionality
 - 85% (-) vs 31% (+)
- Highly significant difference
 - $\chi^2(1, N=99)=27.4, p<0.001$

Aesthetic domain: *neurotypical population*

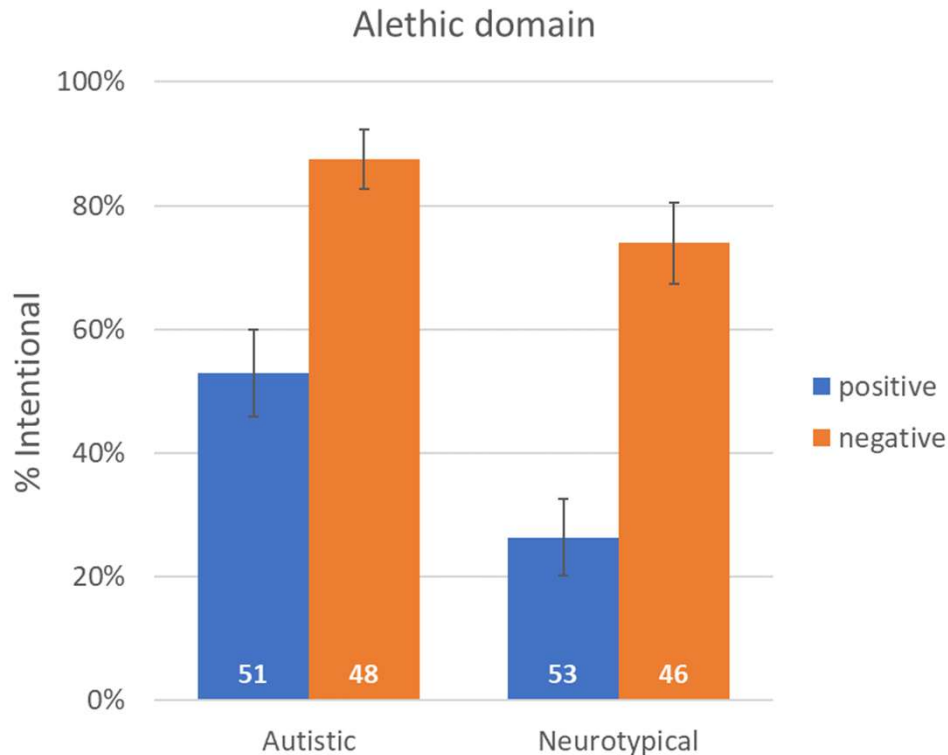
- Results comparable to Knobe's (2004)
 - 65% (-) vs 34% (+)
- Significant difference
 - $\chi^2(1, N=99)=8.4, p=0.0037$

Smaller effect size than in the moral domain

WE ALSO FOUND THE PRESENCE OF THE KNOBE EFFECT IN THE ALETHIC DOMAIN FOR BOTH POPULATIONS

Although the asymmetry is greater in the neurotypical group than in the autistic

Attribution of intentionality (in the alethic domain)



Differences between the positive (+) and negative (-) side-effect cases

Alethic domain: *autistic population*

- Asymmetry in the attribution of intentionality
 - 87% (-) vs 53% (+)
- Highly significant difference
 - $\chi^2(1, N=99)=12.4, p<0.001$

Alethic domain: *neurotypical population*

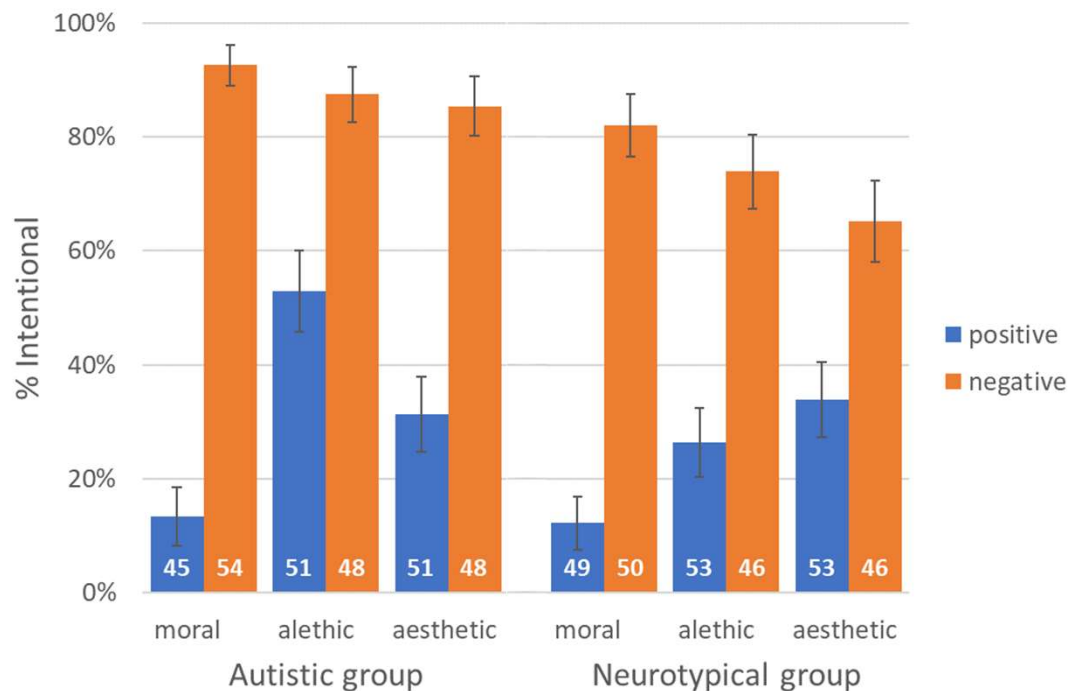
- Asymmetry in the attribution of intentionality
 - 74% (-) vs 26% (+)
- Highly significant difference
 - $\chi^2(1, N=99)=20.4, p<0.001$

Moderate effect size between the moral and the aesthetic domain

THE MAIN DIFFERENCES BETWEEN POPULATIONS ARE FOUND IN TWO SPECIFIC CONDITIONS

Where the intentionality attributed by autistics is higher than that of neurotypicals

Attribution of intentionality (by group and domain)



Main differences between the autistic and neurotypical populations

(1) When the side effect is negative in the aesthetic domain

- Significant difference between populations
 - 85% (autistic) vs 65% (neurotypical)
 - $\chi^2(1, N=94)=4.1, p=0.042$

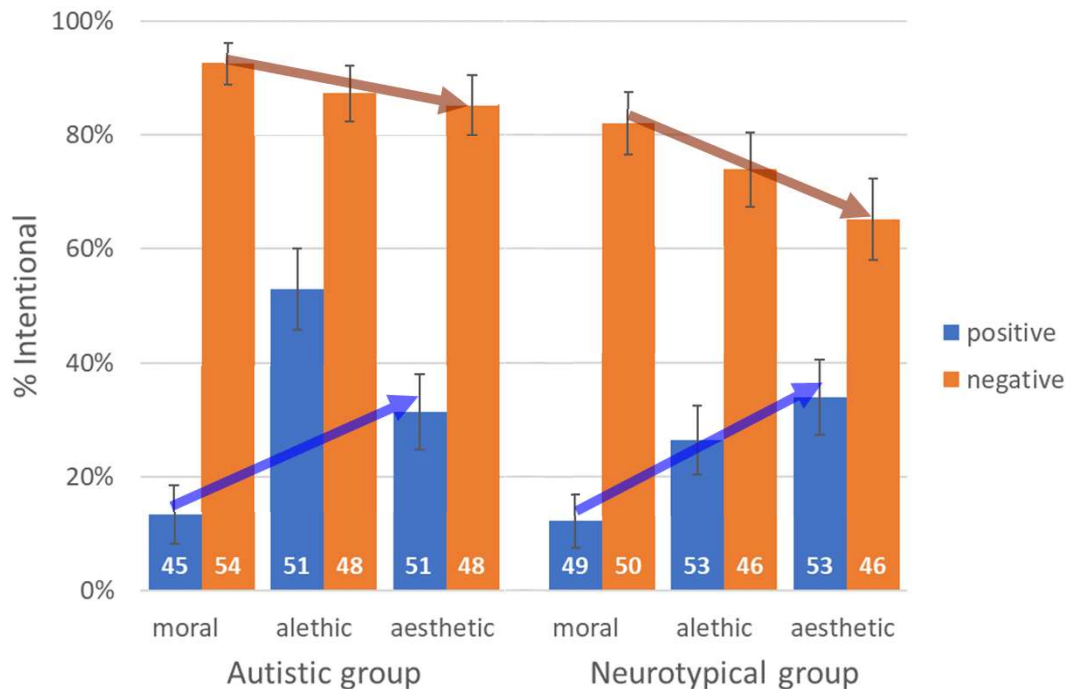
(2) When the side effect is positive in the alethic domain

- Significant difference between populations
 - 53% (autistic) vs 26% (neurotypical)
 - $\chi^2(1, N=104)=6.6, p=0.01$
- This may be due to the theory of mind difficulties exhibited by the autistic group (Baron-Cohen, 2001; Garcia-Molina & Clemente-Estevan, 2019)

A GRADATION CAN BE ESTABLISHED FOR THE THREE DOMAINS IN BOTH POPULATIONS

According to the effect size and the intentionality attributed in the negative cases

Attribution of intentionality (by group and domain)



Comparison of effect sizes⁽¹⁾ by population group and domain

Autistic population

- *Moral* domain: large effect size (VC=0.776)
- *Alethic* domain: medium effect size (VC=0.354)
- *Aesthetic* domain: large effect size (VC=0.526)

Neurotypical population

- *Moral* domain: large effect size (VC=0.678)
- *Alethic* domain: medium effect size (VC=0.435)
- *Aesthetic* domain: small effect size (VC=0.292)

A descending gradation of intentionality attribution for negative side effect cases

moral > alethic > aesthetic

(1) The effect size was determined for the three domains using Cramer's V (VC)

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

- ATTRIBUTION OF INTENTIONALITY
- ATTRIBUTION OF PRAISE AND BLAME

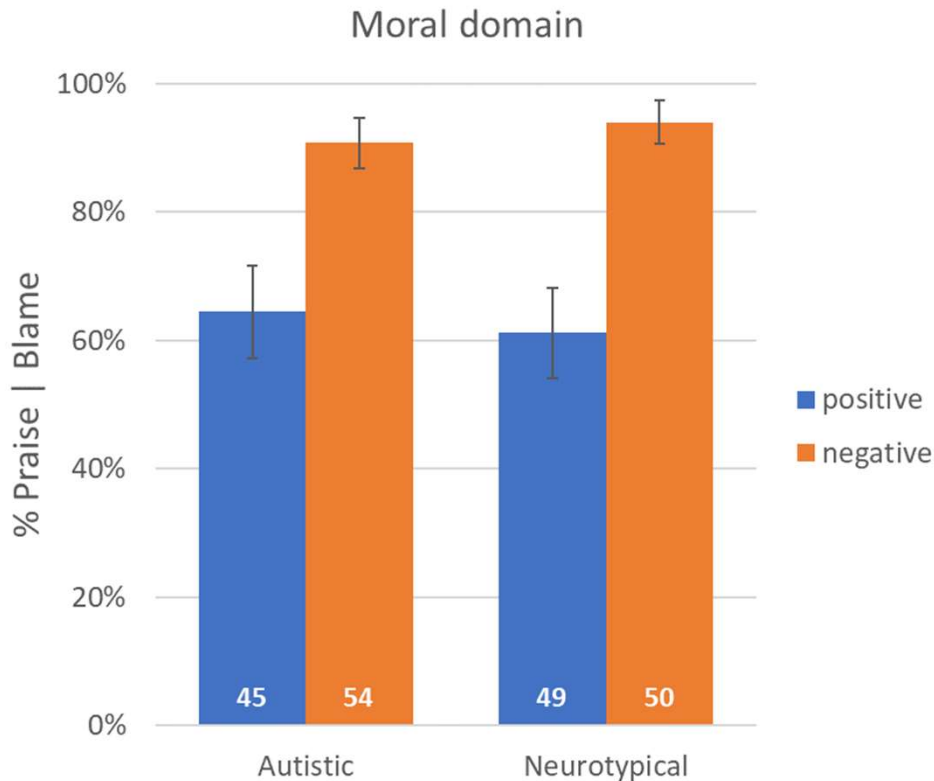
IV. DISCUSSION OF RESULTS

CONCLUSIONS

IN THE MORAL DOMAIN, THE ATTRIBUTION OF BLAME IS HIGHER THAN THE ATTRIBUTION OF PRAISE

Although there are no differences between autistics and neurotypicals

Attribution of praise and blame (in the moral domain)



Differences between the positive (+) and negative (-) side-effect cases

Moral domain: *attribution of blame*

- Percentage who blamed the chairman's action
 - ... autistic (91%) / neurotypical (94%)
- No significant difference
- Blame attribution was very similar to that reported by Zalla & Leboyer (2011)

Moral domain: *attribution of praise*

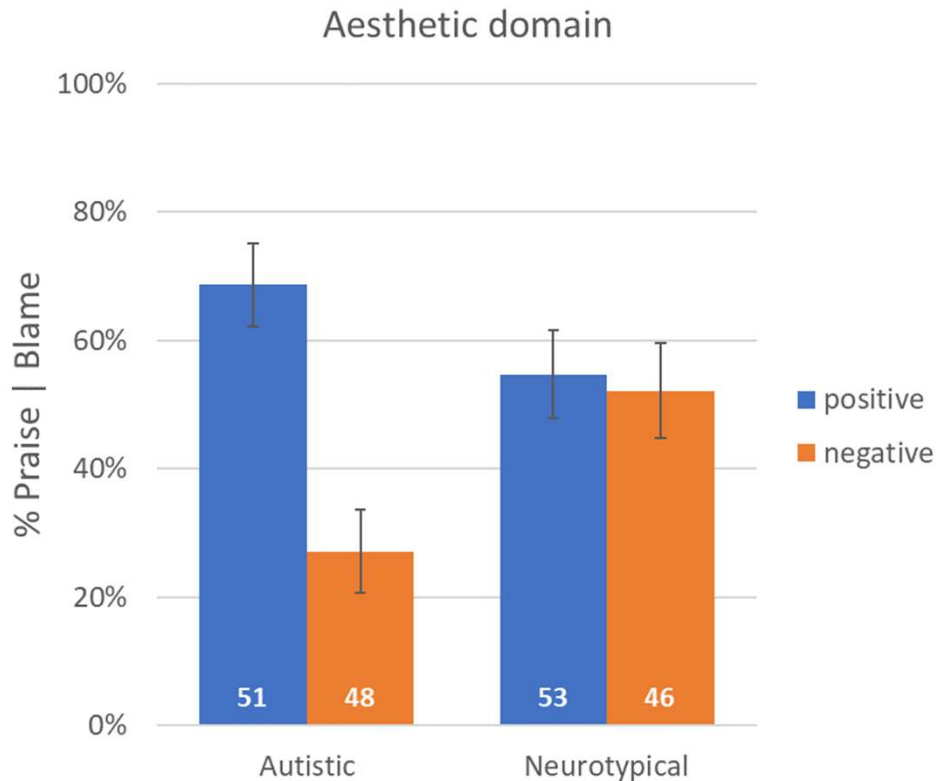
- Percentage who praised the chairman's action
 - ... autistic (64%) / neurotypical (61%)
- No significant difference
- In contrast to Zalla & Leboyer (2011), who found a significant difference⁽¹⁾
 - ... autistic (17%) / neurotypical (43%)

(1) This difference may be due to the much smaller sample size of Zalla and Leboyer's study (N=46 compared to 99 participants in our study)

BLAME ATTRIBUTION WAS SIGNIFICANTLY HIGHER FOR AUTISTICS THAN FOR NEUROTYPICALS IN AESTHETIC CASES

Praise attributions were similar and in line with those of the moral domain

Attribution of praise and blame (in the aesthetic domain)



Differences between the positive (+) and negative (-) side-effect cases

Aesthetic domain: *attribution of blame*

- Percentage who blamed the mayor's decision
 - ... autistic (27%) / neurotypical (52%)
- Significant difference
 - $\chi^2(1, N=94)=5.2, p=0.023$

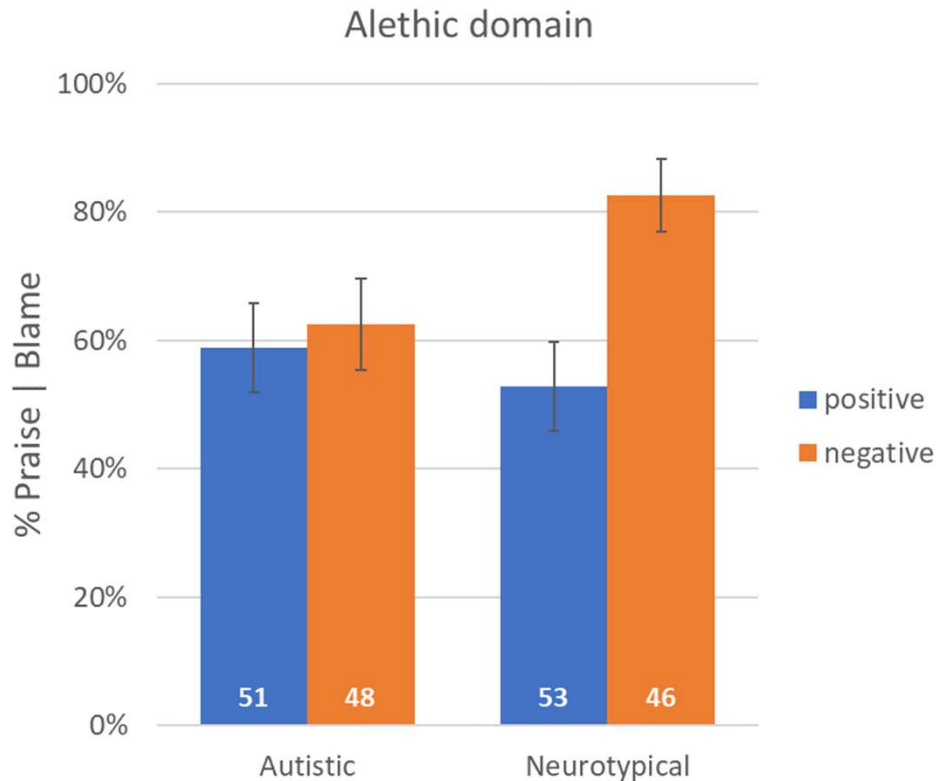
Aesthetic domain: *attribution of praise*

- Percentage who praised the mayor's decision
 - ... autistic (69%) / neurotypical (55%)
- No significant difference

BLAME ATTRIBUTION WAS SIGNIFICANTLY HIGHER FOR NEUROTYPICALS THAN FOR AUTISTICS IN ALETHIC CASES

Praise attributions were similar and in line with the moral and aesthetic domains

Attribution of praise and blame (in the alethic domain)



Differences between the positive (+) and negative (-) side-effect cases

Alethic domain: *attribution of blame*

- Percentage who blamed the candidate
 - ... autistic (62%) / neurotypical (83%)
- Significant difference
 - $\chi^2(1, N=94)=3.8, p=0.05$

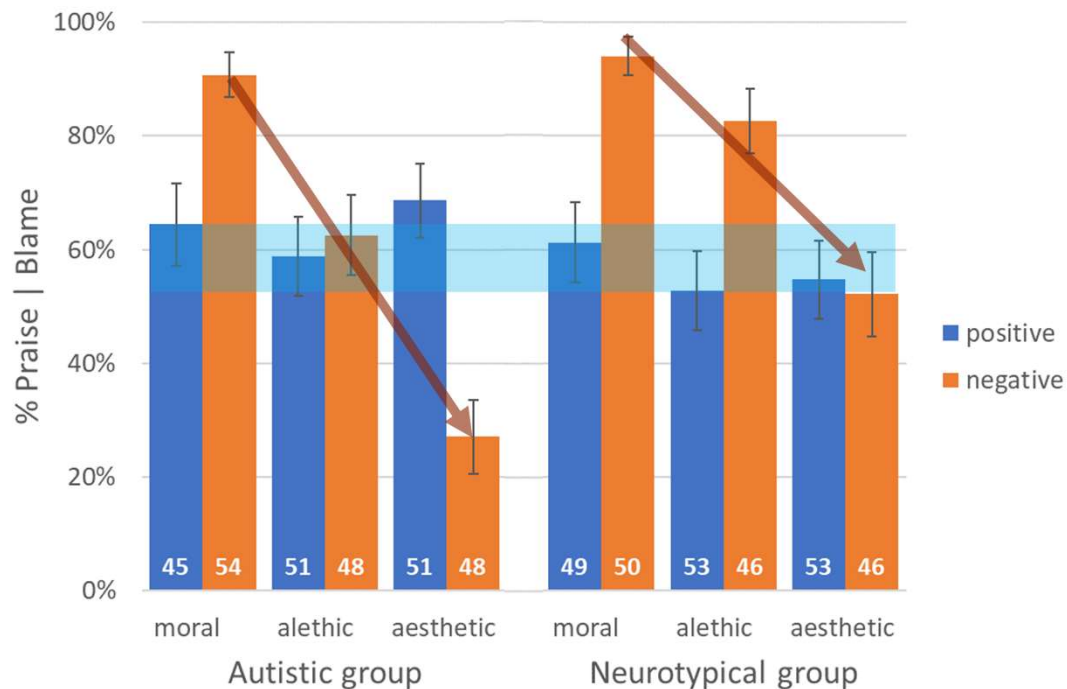
Alethic domain: *attribution of praise*

- Percentage who praised the candidate
 - ... autistic (59%) / neurotypical (53%)
- No significant difference

TWO PATTERNS EMERGE IN THE ATTRIBUTION OF BLAME OR PRAISE FOR THE AGENT'S ACTIONS

Comparison of praise and blame attributions by domain

Attribution of praise and blame (by group and domain)



Patterns identified in the attribution of blame or praise to agents

- (1) For the *positive side effects* ...
... the % of participants attributing praise remains constant at around 60%
 - for both the autistic (~64%)
 - and the neurotypical (~56%) groups
- (2) For the *negative side effects* ...
... there is a downward trend in the attribution of blame with domain
 - *moral* domain → high
 - *alethic* domain → medium
 - *aesthetic* domain → low

DIFFERENCES IN BLAME ATTRIBUTIONS BETWEEN DOMAIN PAIRS BY POPULATION GROUP

Group	Domain pair	χ^2	<i>N</i>	<i>p</i> -value	
Autistic	Moral (91%) vs alethic (62%)	10.0	102	.0015	**
	Moral (91%) vs aesthetic (27%)	40.6	102	<.001	***
	Alethic (62%) vs aesthetic (27%)	10.8	96	.001	***
Neurotypical	Moral (94%) vs alethic (83%)	2.0	96	.15	
	Moral (94%) vs aesthetic (52%)	19.6	96	<.001	***
	Alethic (83%) vs aesthetic (52%)	8.4	92	.0038	**

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

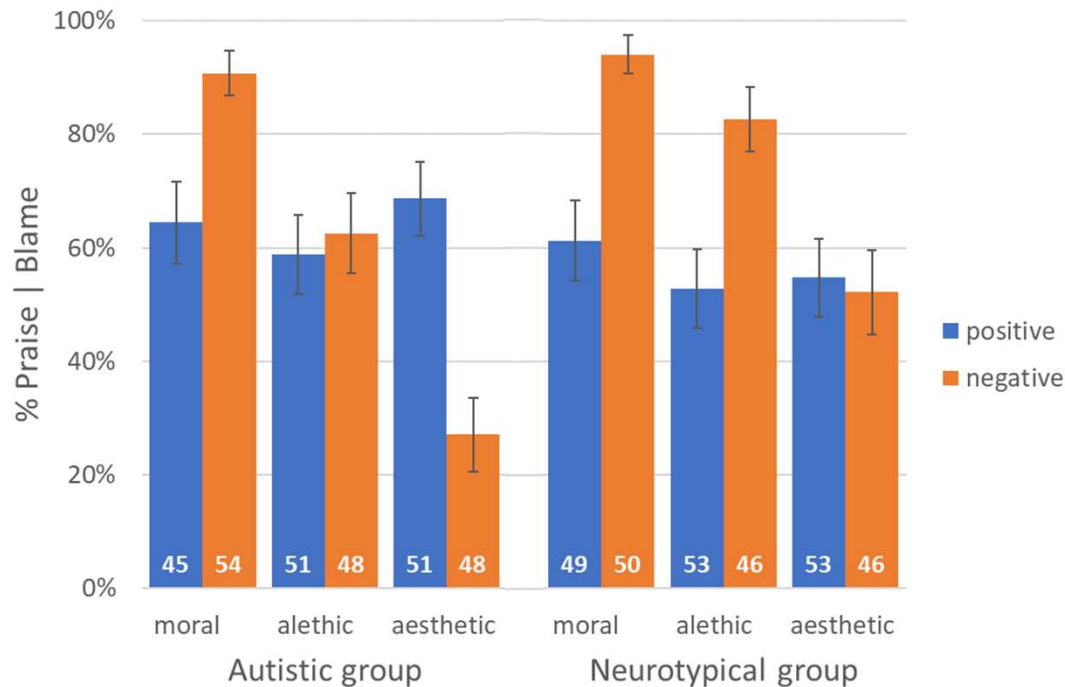
- (1) Differences in attributions of blame are highly significant for all the domain pairs
 - for both autistic and neurotypical populations
 - except the moral vs alethic comparison for neurotypical participants

- (2) In the negative cases, the percentage of autistic people who attribute blame is always lower than the percentage of neurotypical people who do so

THE AMOUNT OF PRAISE AND BLAME ATTRIBUTED TO THE AGENT SHOWS THE SAME TWO PREVIOUS PATTERNS

Representation of attributed praise and blame in the terms used by Knobe (2004)

Attribution of praise and blame (by group and domain)



Amount of praise (+) and blame (-) attributed according to type of side effect

Group	Domain	Praise	Blame
Autistic	Moral	1.71	-2.54
	Alethic	1.51	-1.88
	Aesthetic	1.84	-0.58
Neurotypical	Moral	1.63	-2.62
	Alethic	1.23	-2.11
	Aesthetic	1.42	-1.04

(1) There is a hierarchy of attributed blame common to both populations

- *less* 🤔 causing aesthetic damage
- *more* 😱 inducing false beliefs
- *most* 😱 harming the environment

(2) Neurotypicals always judge cases as more blameworthy than autistics

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

CONCLUSIONS

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

- INTENTIONALITY ATTRIBUTION
- MORAL EVALUATION
- GENERAL COMPARISON OF AUTISTICS VS NEUROTYPICALS
- BACK TO THE KNOBE EFFECT

CONCLUSIONS

PRESENCE OF THE KNOBE EFFECT AND INTERPRETATION

The results with neurotypical and autistic people

- ... confirmed previous findings in the moral (and aesthetic) domains
- ... proved the presence of the Knobe effect also in the alethic domain
- ... showed that the existence of the asymmetry does not depend on the type of damage provoked by the side effect

However, the effect size is not the same in the three domains

- A gradation in the asymmetry of the intentionality attributed by people can be established (in terms of the effect size)
moral > alethic > aesthetic
- The gradation is confirmed by how neurotypicals assigned blame to the agent's decision in negative cases

The amount of praise attributed in positive cases was consistently high

- ... this supports Knobe's idea that intentionality judgments are triggered by consequences
- ... regardless of how praiseworthy or blameworthy actions are judged.

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

- INTENTIONALITY ATTRIBUTION
- MORAL EVALUATION
- GENERAL COMPARISON OF AUTISTICS VS NEUROTYPICALS
- BACK TO THE KNOBE EFFECT

CONCLUSIONS

INTERPRETATION OF PRAISE / BLAME JUDGMENTS (I)

The gradation in the attribution of intentionality was confirmed by the praise and blame that participants attributed to the agents' actions

- No differences between autistic and non-autistic in the moral domain
- There was a different pattern in the alethic and aesthetic domains
 - the autistic evaluation is less negative in cases with *negative effects*
 - and quite similar in cases with *positive effects*

Overall, the results suggest a profile of autistic evaluations consistent with moral judgments based on the agent's declared intentions

- In negative cases autistic people tend –on average– to consider agents less blameworthy than neurotypicals

However, there are reasons to suspect that this is not the case ...

(continues)

INTERPRETATION OF PRAISE / BLAME JUDGMENTS (II)

However, there are reasons to suspect that this is not the case

- (1) For the harming the environment, autistics do not seem to make their moral judgments independent of outcomes
- (2) If judgments were made based on declared intentions, we should not expect differences between cases with good and bad outcomes
- (3) Nevertheless ... in cases with good consequences autistic people give agents a level of praise very similar to neurotypicals

Taken together, the results suggest that

- Autistics are as consequentialist as non-autistics in positive cases
- ... but less consequentialist in cases with negative outcomes
- ❖ Which goes against the view that moral judgments in the autistic group population are more outcome-based due to theory of mind difficulties

INTERPRETATION OF PRAISE / BLAME JUDGMENTS (III)

Nevertheless, it is difficult to hold that autistics blame more based on intentions than based on outcomes, ...

... because in intention attribution

- (1) There is a strong Knobe effect in the autistic group
- (2) (In negative cases) Autistic attributions of intentionality are as high as those of neurotypicals

Finally, our results do not suggest a more ingrained sense of morality in autistic individuals, as

- ... neutral evaluations are also more extended in the autistic group
- ... which suggests that autistics might be more ambivalent than neurotypicals (especially in judging how blameworthy the agent's action is)
- ❖ This is consistent with evidence that higher levels of autism are associated with atypical patterns of moral judgment

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

- INTENTIONALITY ATTRIBUTION
- MORAL EVALUATION
- GENERAL COMPARISON OF AUTISTICS VS NEUROTYPICALS
- BACK TO THE KNOBE EFFECT

CONCLUSIONS

MANY SIMILARITIES AND SOME DIFFERENCES

Main similarities between autistics and neurotypicals

- (1) Strong presence of the Knobe effect in all three domains
- (2) Both groups show a stable hierarchy of evaluations by domain
 - *For negative effects*: subjects attributed more intentionality and blame in the moral domain, then in the alethic, and finally in the aesthetic
 - *For positive effects*: the amount of praise attributed is very similar in all cases (~60%), with no significant differences by domain or group

... and some differences

- (1) Intentionality attribution is higher in autistics (in + and – cases)
 - ... and particularly high (compared to neurotypicals) in the aesthetic negative case and in the alethic positive case
- (2) Autistics praise more (and blame less) in all domains
 - ... exhibiting in those cases a greater difference between intention and blame/praise judgments

THERE IS A CLEAR KNOBE EFFECT COMBINED WITH HIGHER PRAISE THAN BLAME

A phenomenon that is even more pronounced in the autistic population

For instance, autistics seem to think that

- ... the wrong that the mayor brought about is permissible
- ... but –at the same time– that he intended it

This suggests that Knobe's point (i.e., that intention judgments are divorced from moral judgments) is even more acute in the case of autistics

We expected to observe

- (1) Higher moral standards in praise / blame attributions in autistics
 - This difference did not show up
 - (2) A larger Knobe effect in the autistic group
 - This prediction was confirmed in the three domains
- (**) *Hypothesis*: these results could relate to difficulties distinguishing between incidental and intentional outcomes ...
- ... and some persistent differences in means-ends evaluations

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

- INTENTIONALITY ATTRIBUTION
- MORAL EVALUATION
- GENERAL COMPARISON OF AUTISTICS VS NEUROTYPICALS
- BACK TO THE KNOBE EFFECT

CONCLUSIONS

NORMATIVIST INTERPRETATION IS THE MOST EXPLANATORY

Normativist explanation

- The asymmetry would be due to the fact that people form stronger beliefs in cases with negative side effects ...
- ... for being cases of violation of rules

The normativist interpretation can explain the descending rate of intentionality attribution found across domains

- The highest rate happens in the moral domain ...
 - ... where the side effect may be clearly seen as a norm violation of the environmental protection laws
- Attributions of intentionality and blame are lower in the alethic domain ...
 - ... because the induction of wrong beliefs is less clearly associated with a norm violation
- The lowest rate is found in the aesthetic domain
 - ... where the negative side effect was much more difficult to identify as a violation of a norm

AGENDA

AGENDA

I. INTRODUCTION

II. EXPERIMENT DESCRIPTION

III. PRESENTATION OF KEY FINDINGS

IV. DISCUSSION OF RESULTS

CONCLUSIONS

CONCLUSIONS

- ❖ The Knobe effect was present in all domains both in autistic and neurotypical
- ❖ A gradation of domains is possible, in basis of attribution of intentionality (or of blame judgments in negative cases)
Moral > Alethic > Aesthetic
- ❖ Results support Knobe's idea that attributions of intentionality are triggered by the consequences of actions
- ❖ Autistics are more ambivalent in attributing blame
- ❖ Results are more consistent with the normativist interpretation of the Knobe effect

Thanks for listening

FOR QUESTIONS AND COMMENTS: jhercon@uva.es
agustin.vicente@ehu.eus

REFERENCES (I)

- Alfano, M., Beebe, J. R. and Robinson, R. 2012. The centrality of belief and reflection in Knobe-effect cases: A unified account of the data. *The Monist* 95(2): 264-289
- Adams, F. and Steadman, A. 2004. Intentional action in ordinary language: Core concept or pragmatic understanding. *Analysis* 64(2): 173-182
- Baron-Cohen, S. 2001. Theory of mind and autism: A review. In L. M. Glidden (Ed.), *International Review of Research in Mental Retardation: Autism* (Vol. 23, pp. 169-184). New York: Academic Press
- Beebe, J. R. 2013. A Knobe effect for belief ascriptions. *Review of Philosophy and Psychology* 4: 235-258
- Beebe, J. R., and Buckwalter, W. 2010. The epistemic side-effect effect. *Mind and Language* 25(4): 474-498
- Cohen-Eliya, M. and Porat, I. 2015. The Knobe effect, indifference, and constitutional law. *Law & Ethics of Human Rights* 9(2): 229-247
- Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., Beebe, J., Berniūnas, R., Boudesseul, J., Colombo, M., Cushman, F., Díaz, R., van Dongen, N. N. N., Dranseika, V., Earp, B. D., Gaitán Torres, A., Hannikainen, I., Hernández-Conde, J. V., Hu, W., Jaquet, F., Khalifa, K., Kim, H., Kneer, M., Knobe, J., Kurthy, M., Lantian, A., Liao, S., Machery, E., Moerenhout, T., Mott, C., Phelan, M., Phillips, J., Rambharose, N., Reuter, K., Romero, F., Sousa, P., Sprenger, J., Thalabard, E., Tobia, K., Viciano, H., Wilkenfeld, D. and Zhou, X. 2021. Estimating the reproducibility of experimental philosophy. *Review of Philosophy and Psychology* 12: 9-44
- Dalbauer, N. and Hergovich, A. 2013. Is what is worse more likely? The probabilistic explanation of the epistemic side-effect effect. *Review of Philosophy and Psychology* 4: 639-657
- Dempsey, E. E., Moore, C., Johnson, A., Stewart, S. H., and Smith, I. M. (2020). Morality in autism spectrum disorder: A systematic review. *Development and Psychopathology* 32(3): 1069-1085
- Garcia-Molina, I., and Clemente-Estevan, R. A. (2019). Autism and faux pas. Influences of presentation modality and working memory. *The Spanish Journal of Psychology* 22, e13: 1-11

REFERENCES (II)

- Harman, G. 1976. Practical reasoning. *The Review of Metaphysics* 29(3): 431-463
- Holton, R. 2010. Norms and the Knobe effect. *Analysis* 70(3): 417-424
- Hu, Y., Pereira, A. M., Gao, X., Campos, B. M., Derrington, E., Corgnet, B., Zhou, X., Cendes, F., and Dreher, J.-C. (2021). Right temporoparietal junction underlies avoidance of moral transgression in autism spectrum disorder. *The Journal of Neuroscience* 41(8): 1699-1715
- Kaspar, K., Newen, A., Dratsch, T., de Bruin, L., Al-Issa, A. and Bente, G. 2016. Whom to blame and whom to praise. *International Journal of Cross Cultural Management* 16(3): 341-365
- Knobe, J. 2003a. Intentional action and side effects in ordinary language. *Analysis* 63(3): 190-194
- . 2003b. Intentional action in folk psychology: An empirical investigation. *Philosophical Psychology* 16(2): 309-324
- Knobe, J. 2004. Folk psychology and folk morality: Response to critics. *Journal of Theoretical and Philosophical Psychology* 24(2): 270-279
- Knobe, J., and Burra, A. 2006. The folk concepts of intention and intentional action: A cross-cultural study. *Journal of Cognition and Culture* 6(1-2): 113-132
- Leslie, A. M., Knobe, J., and Cohen, A. 2006. Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science* 17: 421-427
- Machery, E. 2008. The folk concept of intentional action: Philosophical and experimental issues. *Mind and Language* 23(2): 165-189
- Margoni, F., and Surian, L. (2016). Mental state understanding and moral judgment in children with autistic spectrum disorder. *Frontiers in Psychology* 7: 1478
- Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., and Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences* 108(7): 2688-2692

REFERENCES (III)

- Nadelhoffer, T. 2004. Blame, badness, and intentional action: A reply to Knobe and Mendlow. *Journal of Theoretical and Philosophical Psychology* 24(2): 259-269
- . 2006. Bad acts, blameworthy agents, and intentional actions: Some problems of juror impartiality. *Philosophical Explorations* 9(2): 203-219
- Nichols, S. and Ulatowski, J. 2007. Intuitions and individual differences: The Knobe effect revisited. *Mind and Language* 22(4): 346-365
- Pellizzoni, S., Siegal, M., and Surian, L. 2009. Foreknowledge, caring, and the side-effect effect in young children. *Developmental Psychology* 45(1), 289-295
- Petrolini, V., Jorba, M., and Vicente, A. (2023). What does it take to be rigid? Reflections on the notion of rigidity in autism. *Frontiers in Psychiatry* 14: 1072362
- Pettit, D. and Knobe, J. 2009. The pervasive impact of moral judgment. *Mind and Language* 24(5): 586-604
- Phelan, M. and Sarkissian, H. 2009. Is the 'trade-off hypothesis' worth trading for? *Mind and Language* 24(2): 164-180
- Shulman, C., Guberman, A., and Shiling, N. (2012). Moral and social reasoning in autism spectrum disorders. *Journal of Autism and Developmental Disorders* 42(7): 1364-1376
- Utikal, V. and Fischbacher, U. 2014. Attribution of externalities: An economic approach to the Knobe effect. *Economics and Philosophy* 30(2): 215-240
- Wible, A. 2009. Knobe, side effects, and the morally good business. *Journal of Business Ethics* 85(1): 173-178
- Young, L., Cushman, F., Adolphs, R., Tranel, D. and Hauser, M. 2006. Does emotion mediate the relationship between an action's moral status and its intentional status? Neuropsychological evidence. *Journal of Cognition and Culture* 6(1-2): 291-304
- Zalla, T. and Leboyer, M. 2011. Judgment of intentionality and moral evaluations in individuals with High Functioning Autism. *Review of Philosophy and Psychology* 2: 681-698