Contextual inferences and conversation profiles: a comparison between autistic and neurotypical adults

Elena Castroviejo
(Joint work with Agustín Vicente, Mark Jary,
and Isabel Martín-González)

UPV/EHU & Lindy Lab

KEAMA Workshop

Universidad de Valladolid, March 7, 2024

Laying the ground

◎ Individuals on the autism spectrum are characterized by clinicians as having difficulties with pragmatic abilities, irrespective of structural language skills, severity, IQ or age.

◎ Very broad notions of **pragmatics** in psychological assessment and research. For example: respecting speech turns, understanding non-literal language, gaze behavior.

◎ We were interested in making finer-grained descriptions of these pragmatic skills on the basis of theoretical knowledge in **formal** semantics and pragmatics.

Laying the ground

◎ Individuals on the autism spectrum are characterized by clinicians as having difficulties with pragmatic abilities, irrespective of structural language skills, severity, IQ or age.

◎ Very broad notions of **pragmatics** in psychological assessment and research. For example: respecting speech turns, understanding non-literal language, gaze behavior.

◎ We were interested in making finer-grained descriptions of these pragmatic skills on the basis of theoretical knowledge in **formal** semantics and pragmatics.

Laying the ground

◎ Individuals on the autism spectrum are characterized by clinicians as having difficulties with pragmatic abilities, irrespective of structural language skills, severity, IQ or age.

◎ Very broad notions of **pragmatics** in psychological assessment and research. For example: respecting speech turns, understanding non-literal language, gaze behavior.

◎ We were interested in making finer-grained descriptions of these pragmatic skills on the basis of theoretical knowledge in **formal** semantics and pragmatics.

Laying the ground

Conversational skills (Ying Sng et al. 2018; Ramos-Cabo et al., submitted)

◎ Difficulties in engaging in conversation as related to ToM (Hale and Tager-Flusberg 2005). Especially: Topic maintenance, initiations, relevance of contributions.

◎ Mixed results in the recent literature, especially due to to the diversity of annotation protocols and loose definitions of studied variables.

◎ Potential effect of non-linguistic variables: mixed vs. non-mixed pairs (Crompton et al. 2020, Bauminger-Zviely et al. 2014, 2017).

Laying the ground

Conversational skills (Ying Sng et al. 2018; Ramos-Cabo et al., submitted)

◎ Difficulties in engaging in conversation as related to ToM (Hale and Tager-Flusberg 2005). Especially: Topic maintenance, initiations, relevance of contributions.

◎ Mixed results in the recent literature, especially due to to the diversity of annotation protocols and loose definitions of studied variables.

◎ Potential effect of non-linguistic variables: mixed vs. non-mixed pairs (Crompton et al. 2020, Bauminger-Zviely et al. 2014, 2017).

Laying the ground

Conversational skills (Ying Sng et al. 2018; Ramos-Cabo et al., submitted)

◎ Difficulties in engaging in conversation as related to ToM (Hale and Tager-Flusberg 2005). Especially: Topic maintenance, initiations, relevance of contributions.

◎ Mixed results in the recent literature, especially due to to the diversity of annotation protocols and loose definitions of studied variables.

◎ Potential effect of non-linguistic variables: mixed vs. non-mixed pairs (Crompton et al. 2020, Bauminger-Zviely et al. 2014, 2017).

This talk

◎ Building on the work by Wilson and Bishop (2019, 2020, 2022), we address the question of whether the apparent difficulties autistic people have shown with conversational implicature could really be due to issues with global/central coherence (**Experiment 1**).

◎ Report a follow-up experiment that aims at addressing the question of whether there is a different conversational profile in autism (**Experiment 2**).

This talk

◎ Building on the work by Wilson and Bishop (2019, 2020, 2022), we address the question of whether the apparent difficulties autistic people have shown with conversational implicature could really be due to issues with global/central coherence (**Experiment 1**).

◎ Report a follow-up experiment that aims at addressing the question of whether there is a different conversational profile in autism (**Experiment 2**).

Experiment 1:
Is it implicature or local processing?

Background: Wilson & Bishop (WB)

Through the **Implicature Comprehension Test**, WB (2019, 2020, 2021, 2022), observe that

- ◎ Autistic individuals (AUT) are more likely to choose a "non-normative" interpretation of an implied meaning than Neurotypicals (NT) (accuracy of 62% vs. 80%).

- ◎ **Core language** difficulties do not explain inferencing difficulties.

(1)   S: Can the two of us sit here?
        T: The children just went to find the toilet.
              Do you think T and S can sit there? ⇒ No.

Background: Wilson & Bishop (WB)

Through the **Implicature Comprehension Test**, WB (2019, 2020, 2021, 2022), observe that

- ◎ Autistic individuals (AUT) are more likely to choose a "non-normative" interpretation of an implied meaning than Neurotypicals (NT) (accuracy of 62% vs. 80%).

- ◎ **Core language** difficulties do not explain inferencing difficulties.

(2)   S: The seats here are all taken.
         T: The carriage behind is much less busy.
               Do you think S and T will find a seat?   ⇒ Yes.

Changes to the paradigm

◎ We adapt WB's items to create a **story** to make it more attractive.

◎ In their papers, WB suggest that it is difficult to tell whether there is a problem with implicature or with local processing.

   ⊕ We include Local processing (called **Context**) as a separate condition from Implicature.

   ⇒ From now on, I will only focus on the Context condition.

◎ We do not have a direct prompting question, but rather ask the participant to choose the preferred **continuation** of the story.

Changes to the paradigm

◎ We adapt WB's items to create a **story** to make it more attractive.

◎ In their papers, WB suggest that it is difficult to tell whether there is a problem with implicature or with local processing.

⊕ We include Local processing (called **Context**) as a separate condition from Implicature.

⇒ From now on, I will only focus on the Context condition.

◎ We do not have a direct prompting question, but rather ask the participant to choose the preferred **continuation** of the story.

Changes to the paradigm

◎ We adapt WB's items to create a **story** to make it more attractive.
◎ In their papers, WB suggest that it is difficult to tell whether there is a problem with implicature or with local processing.
  ⊕ We include Local processing (called **Context**) as a separate condition from Implicature.
  ⇒ From now on, I will only focus on the Context condition.
◎ We do not have a direct prompting question, but rather ask the participant to choose the preferred **continuation** of the story.

Participants

A total of 38 individuals (Spanish-speaking)

◎ AUT (N = 19, ages 14–54, mean age 30.2)

⊕ Verbal, IQ within typical range.

⊕ Recruited from our pool of participants in Vitoria-Gasteiz + online.

◎ NT (N = 19, ages 13-56, mean age 26.1)

⊕ Recruited through acquaintances, not university students.

Design

- ◎ (In Spanish)
- ◎ A **forced-choice task** with 3 possible options (participant has to choose the **BEST** possible answer).
- ◎ Each trial is a **dialogue** between two characters (Pedro and Carmen), embedded within a **story**.
- ◎ Format of the trial:
    1. Sentence describing the context.
    2. Dialogue (utterance 1, utterance 2).
    3. Prompting question: How does the conversation continue?
    4. 3 options: A, B, C.

Design

- ◎ x 16 experimental items, x 16 filler items (only 1 obvious good answer: x 8 of type A, x 8 of type B)
- ◎ Independent variable (between subjects): group (NT vs. AUT)
- ◎ Dependent variable: type of answer
    1. Condition **A**: answer addresses the global dialogue (Utterance 1 + Utterance 2).
    2. Condition **B**: answer addresses Utterance 2, but not Utterance 1.
    3. Condition **C**: random answer.
- ◎ Pseudo-randomization of items, all participants saw them in the same order (to preserve the logic of the narrative) + pseudo-randomization of type of answer in each trial.

Design

Interpretation of the answer type for the critical items:

Type A	This answer elaborates on Utterance 2's reply to the Question Under Discussion (Roberts, 1996) raised by Utterance 1. It evaluates U2, elaborates on the consequences of U2 or addresses a "Do you also p?" implicit sub-question (**topic maintenance**). As such, it observes the overall context and involves global coherence. **PREFERRED ANSWER**.

Design

Interpretation of the answer time for the critical items (ctd):

Type B	This answer exclusively addresses a "What about x?" sub-question, building on one of the constituents of U2, which is otherwise disconnected of the QUD raised by U1. As such, there is a topical relation with U2 alone, and it is a legitimate **topic shift** (Van Kuppevelt, 1995). Choosing this answer involves observing locality, but not taking into account the issue raised by U1 and, thus, global coherence. **LESS PREFERRED GOOD ANSWER**.
Type C	This is an answer that does not observe any discourse logic. **WRONG ANSWER**.

Design

Trial example:

(3)   [Context]: It's only been a short while, but Pedro is already quite
        bored. He doesn't like long trips.

        Pedro:Is the train going to stop many more times?
        Carmen:Don't worry; it only has two stops.
            How does the conversation continue?

(4)   Pedro:Ah, OK, it's not much.   TYPE A
        Pedro:On the next stop, many people will get on.   TYPE B
   Pedro:Streetcars are electricity-driven.   TYPE C

Hypotheses

1. If there is an issue with global coherence in AUT $\Rightarrow$ AUT participants may opt for Type B answers than NTs.
2. If AUT have a different conversation profile than NT $\Rightarrow$ Type B and C answers will be chosen more often in AUT than in NT.

Hypotheses

1. If there is an issue with global coherence in AUT $\Rightarrow$ AUT participants may opt for Type B answers than NTs.

2. If AUT have a different conversation profile than NT $\Rightarrow$ Type B and C answers will be chosen more often in AUT than in NT.

Results

Results by Response Type and Group (N=38)
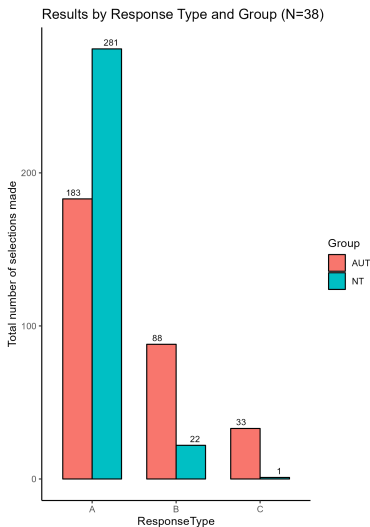
◎ NTs opt more often for Type A than AUTs.

◎ NTs mostly choose the B type answer when they fail, whereas AUTs sometimes prefer the C type answer.
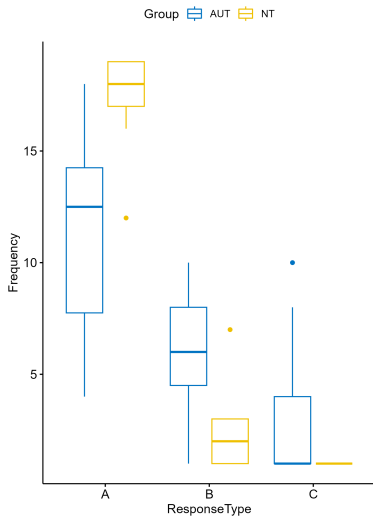
Results

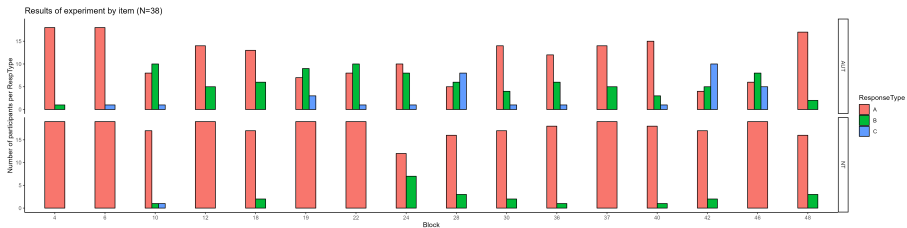**Figure:** Boxplot by group and response type

Results

Figure: Number of participants per RespType for each item

Analysis

- ◎ Mixed effects model with random effects (participant and item) without random slopes, and a binary dependent variable (Accuracy: A vs. non-A answer), with Group (AUT, NT) as explanatory variable.

  (5)   ACC ~ Group + Duration + Age_ group (1 | Subject) + (1 | Block)

- ◎ **Main effect of Group**, with accuracy being more likely in the NT group than in the AUT group ($\beta = 2.55$, SE $= 0.36$, Z(608) $= 6.9$, p <0.0001).
  - ⊕ There is a probability of 0.96 for NTs to opt for Type A, whereas it is 0.7 for AUTs. This is a significant difference.

- ◎ Also no effect of duration (second parts are slightly less accurate than first parts, but p $= 0.5$), and a subtle effect of age group (younger participants are less accurate, p $= 0.09$).

Discussion

◎ There is a majority of Type A answers overall, but it is true that AUTs, as compared to NTs, choose Type B answers more often. This might be viewed as an issue with **global coherence**.

◎ While Type B answers are the typical alternative to Type A in NTs, AUTs chose Type C quite often. This raises the question of whether there is an **AUT conversation profile** and/or whether some AUTs were not systematically following a pattern.

Post-hoc qualitative analysis

Item #10 (A: 42.1%, **B: 52.63%**, C: 5.26%)

(6)   [Context: El tren se ha puesto en marcha, y Carmen y Pedro van charlando.]

Carmen:¿Tienes ganas de este fin de semana?
Pedro:Sí, me apetece mucho ir a esquiar.
        How does the conversation continue?

(7)   Carmen:Qué bien, lo vamos a pasar fenomenal.   Type A
       Carmen:A mí me gusta más el esquí de fondo que el esquí alpino.
                                                       Type B
       Carmen:En la montaña hace frío.   Type C

Post-hoc qualitative analysis

Item #19 (A: 36.84%, **B: 47.36%**, C: 15.79%)

(8)   [Context: Como Pedro no para de hacerle preguntas, Carmen
       decide dejar de leer y seguir hablando con Pedro. Empiezan a hacer
       planes sobre el fin de semana.]

       Carmen:¿Te apetece tomar un chocolate caliente este fin de
           semana?
       Pedro:Teniendo en cuenta que estamos de vacaciones, quizá mejor
           que el café . . .
           How does the conversation continue?

(9)   Carmen:Estupendo, así desayunamos los dos chocolate.   TYPE A
       Carmen:Es que el café es estimulante.   TYPE B
       Carmen:Tendremos que abrir las ventanas para ventilar.   TYPE C

Post-hoc qualitative analysis

Item #22 (A: 42.1%, **B: 52.63%**, C: 5.26%)

(10)   [Context: Tras bajarse del tren, Carmen busca la ubicación del hotel que han reservado en Google Maps.]

Carmen:¿Te acuerdas del nombre del hotel?
Pedro:No, pero estaba en la calle principal.
   How does the conversation continue?

(11)   Carmen:¡Es verdad! Se llamaba como la calle.   Type A
Carmen:En la calle principal seguro que hay mucho tráfico.

                                                             Type B

Carmen:Cada vez va más gente al fútbol femenino.   Type C

Post-hoc qualitative analysis

Item #28 (A: 26.3%, B: 31.6%, **C: 42.1%**)

(12) [Context: Pedro empieza a deshacer la maleta y meter ropa en cajones.]
      Pedro:¿Dónde has dejado tú el abrigo?
      Carmen:En el armario.
          How does the conversation continue?

(13) Pedro:Ahí está muy bien.   Type A
      Pedro:Es un armario muy grande.   Type B
      Pedro:Me duele la espalda.   Type C

Limitations

Analyzing the results by item. . .

- ◎ We detect inconsistencies among items
    - ⊕ Some Context items should be resolved by generating an implicature.
    - ⊕ Lack of balance between naturalness of A, B and C-type answers.
    - ⊕ Definition of response type too coarse-grained.
- ◎ We detect a plausible reason for the different profile that does not have to do with the divide we had created, but is rather due to speech act type.
    - ⊕ AUTs prefer "action" to "acknowledgment" speech acts.

Experiment 2:
Different conversational profiles?

Hypotheses

The results found in Exp1 may be due to differences in conversational profiles between AUTs and NTs.

1 NTs will opt for Acknowledgments more often than AUTs.
   ⊕ Acknowledgment = assertion that confirms that the question raised has been resolved. **Pro-social move**.

2 AUTs will be more likely to choose a Topic Shift than NTs.
   ⊕ Topic Shift = assertion that departs from the initial strategy-QUD (Roberts, 1996/2012). **Efficiency-driven move**.

Hypotheses

The results found in Exp1 may be due to differences in conversational profiles between AUTs and NTs.

1. NTs will opt for Acknowledgments more often than AUTs.
   - ⊕ Acknowledgment = assertion that confirms that the question raised has been resolved. **Pro-social move**.
2. AUTs will be more likely to choose a Topic Shift than NTs.
   - ⊕ Topic Shift = assertion that departs from the initial strategy-QUD (Roberts, 1996/2012). **Efficiency-driven move**.

Method

- ◎ Participants: 80 NT and 80 AUT adults (total **N = 160**), ages 20-35, British English speakers, recruited from **Prolific**.
- ◎ Total number of trials: 32 (x 16 critical, x 16 fillers, randomized).
- ◎ A forced-choice task with 4 possible options (dependent variable).
    - ⊕ Type **A**: Acknowledgment.
    - ⊕ Type **B**: Associated Topic Shift (assertion that deviates from QUD by elaborating on a DT).
    - ⊕ Type **C**: Implicit Non-Associated TS (assertion that opens a new unrelated QUD, change of DT).
    - ⊕ Type **D**: Generic statement (control).
- ◎ Manipulated variables:
    - ⊕ Between subjects: status (AUT, NT).
- ◎ Each trial is a dialogue between two characters (You and Alex).

Method

Example of a critical trial:

(14)   You: Have you turned up the heating?
        Alex: Maria switched it up high just now.

You respond with one of the following:

(15)   You: It'll soon warm up in here.   ACKNOWLEDGMENT
        You: It's Maria's birthday tomorrow.   ASSOCIATED TS
        You: I'm going to watch a film on TV later this evening.

                                                           IMPLICIT NON-ASS TS
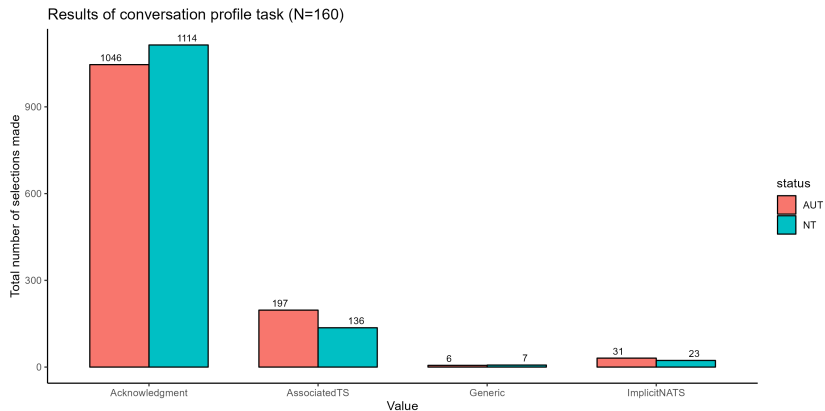   You: Television is a modern invention.   GENERIC

Results

Figure: Number of selections made by Response Type, and Group

Results

Table: Percentage of choices by status

	AUT	NT
Acknowledgment	1046 (81.7%)	1114 (87%)
Associated TS	197 (15.4%)	136 (10.62%)
Implicit Non-associated TS	31 (2.42%)	23 (1.8%)
Generic	6 (0.46%)	7 (0.54%)
Total	1280 (100%)	1280 (100%)

Results

Table: Mean and SD by item

	AUT	NT
Acknowledgment	65.4 (9.64)	69.6 (7.65)
Associated TS	12.3 (9.18)	8.5 (7.47)
Implicit Non-associated TS	2.82 (2.27)	2.09 (1.22)
Generic	1.2 (0.44)	1 (0)

Results
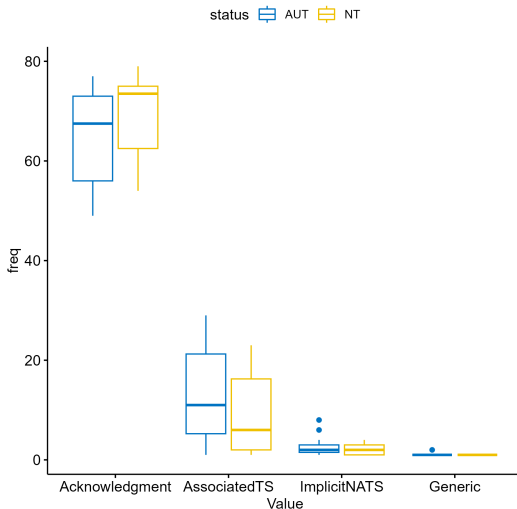
Figure: Distribution of responses by item in the two groups

Results

⊚ We recoded our dependent variable as a dichotomical variable ($1 =$ Acknowledgment, $0 =$ any other option) to fit a mixed effects logistic regression.

(16)   Accuracy $\sim$ status $+ (1 \mid$ item$) + (1 \mid$ participant$)$

⊕ **Main effect of Group**, with Accuracy being more likely in the NT group than in the AUT group ($\beta = 0.47$, SE $= 0.19$, Z(2560) $= 2.36$, p $= 0.0179$)

⊕ There is a probability of 0.92 for NTs to opt for Acknowledgment, whereas it is 0.88 for AUTs. This is a significant difference.

Discussion and Limitations

◎ AUT participants make very similar choices than NT participants,
clearly favoring the Acknowledgment option. We can't really talk
about an **autistic conversational profile** that builds on our proposed
divide, or else there is no lack of comprehension of NT rules of
conversation.

  ⊕ This is a controlled comprehension experiment, not spontaneous
  speech. It taps into **competence**.

  ⊕ We do not have cognitive or linguistic data on the AUT participants,
  who may have a different cognitive profile than the ones tested in
  Exp1. Note the very few Implicit Non-associated TS and Generics.

◎ **Yet** we have encountered a difference between the choice of
Acknowledgment and the rest of options altogether. It looks as if
AUT participants find it less problematic to opt for a Topic Shift than
NTs (**more tolerant to TS**).

Discussion and Limitations

- ◎ AUT participants make very similar choices than NT participants, **clearly favoring the Acknowledgment option**. We can't really talk about an **autistic conversational profile** that builds on our proposed divide, or else there is no lack of comprehension of NT rules of conversation.
    - ⊕ This is a controlled comprehension experiment, not spontaneous speech. It taps into **competence**.
    - ⊕ We do not have cognitive or linguistic data on the AUT participants, who may have a different cognitive profile than the ones tested in Exp1. Note the very few Implicit Non-associated TS and Generics.
- ◎ **Yet** we have encountered a difference between the choice of Acknowledgment and the rest of options altogether. It looks as if AUT participants find it less problematic to opt for a Topic Shift than NTs (**more tolerant to TS**).

General Discussion

◎ Exp2 yields a picture that is somewhat comparable to Exp1 regarding conversational profiles. While in Exp2 we don't see striking differences, there seems to be a pattern of more tolerance towards Topic Shifts in AUT.

◎ Differences in distribution may be due to differences in the segment of the spectrum that we investigated in each Exp + how response types were defined.

◎ Overall, we observe an intact capacity for **understanding** NT conversation rules. Differences may arise in less controlled settings, notably in spontaneous interactions.

General Discussion

◎ Exp2 yields a picture that is somewhat comparable to Exp1 regarding conversational profiles. While in Exp2 we don't see striking differences, there seems to be a pattern of more tolerance towards Topic Shifts in AUT.

◎ Differences in distribution may be due to differences in the segment of the spectrum that we investigated in each Exp + how response types were defined.

◎ Overall, we observe an intact capacity for **understanding** NT conversation rules. Differences may arise in less controlled settings, notably in spontaneous interactions.

General Discussion

◎ Exp2 yields a picture that is somewhat comparable to Exp1 regarding
conversational profiles. While in Exp2 we don't see striking
differences, there seems to be a pattern of more tolerance towards
Topic Shifts in AUT.

◎ Differences in distribution may be due to differences in the segment of
the spectrum that we investigated in each Exp + how response types
were defined.

◎ Overall, we observe an intact capacity for **understanding** NT
conversation rules. Differences may arise in **less controlled settings**,
notably in spontaneous interactions.

General Discussion

Questions for discussion and future research:

◎ How is this competence acquired in AUT from childhood? What are the cognitive and/or linguistic factors that play a role in the development of this competence?

◎ Do we expect different outcomes on different definitions of response types?

◎ Can we expect more from this paradigm? (I.e., does it make sense to pursue these research questions on the basis of forced-choice tasks?)

◎ Don't these results suggest that conversation structure should/could be trained in intervention?

General Discussion

Questions for discussion and future research:

- ◎ How is this competence acquired in AUT from childhood? What are the cognitive and/or linguistic factors that play a role in the development of this competence?

- ◎ Do we expect different outcomes on different definitions of response types?

- ◎ Can we expect more from this paradigm? (I.e., does it make sense to pursue these research questions on the basis of forced-choice tasks?)

- ◎ Don't these results suggest that conversation structure should/could be trained in intervention?

General Discussion

Questions for discussion and future research:

- ◎ How is this competence acquired in AUT from childhood? What are the cognitive and/or linguistic factors that play a role in the development of this competence?

- ◎ Do we expect different outcomes on different definitions of response types?

- ◎ Can we expect more from this paradigm? (I.e., does it make sense to pursue these research questions on the basis of forced-choice tasks?)

- ◎ Don't these results suggest that conversation structure should/could be trained in intervention?

General Discussion

Questions for discussion and future research:

- ◎ How is this competence acquired in AUT from childhood? What are the cognitive and/or linguistic factors that play a role in the development of this competence?

- ◎ Do we expect different outcomes on different definitions of response types?

- ◎ Can we expect more from this paradigm? (I.e., does it make sense to pursue these research questions on the basis of forced-choice tasks?)

- ◎ Don't these results suggest that conversation structure should/could be trained in intervention?

LINDY LAB
LANGUAGE IN NEURODIVERSITY

Contextual inferences and conversation profiles: a comparison between autistic and neurotypical adults

Thanks for your feedback!

Elena Castroviejo

http://www.lindy-lab.eus

elena.castroviejo@ehu.eus

References I

Bauminger-Zviely, N., Golan-Itshaky, A. and Tubul-Lavy, G.: 2017, Speech acts during friends' and non-friends' spontaneous conversations in preschool dyads with high-functioning autism spectrum disorder versus typical development, *Journal of Autism and Developmental Disorders* **47**(5), 1380–1390.

Bauminger-Zviely, N., Karin, E., Kimhi, Y. and Agam-Ben-Artzi, G.: 2014, Spontaneous peer conversations in preschoolers with high-functioning autism spectrum disorder versus typical development, *Journal of Child Psychology and Psychiatry* **55**(4), 363–373.

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G. and Fletcher-Watson, S.: 2020, Autistic peer-to-peer information transfer is highly effective, *Autism* pp. 1–9.

Hale, C. M. and Tager-Flusberg, H.: 2005, Social communication in children with autism: The relationship between theory of mind and discourse development, *Autism* **9**(2), 157–178.

Wilson, A. and Bishop, D.: 2019, "if you catch my drift...": Ability to infer implied meaning is distinct from vocabulary and grammar skills, *Wellcome Open Research* **4**, 68.

References II

Wilson, A. C. and Bishop, D. V. M.: 2020, Judging meaning: A domain-level difference between autistic and non-autistic adults, *Royal Society Open Science* **7**(11), 200845.

Wilson, A. C. and Bishop, D. V. M.: 2022, A novel online assessment of pragmatic and core language skills: An attempt to tease apart language domains in children, *Journal of Child Language* **49**(1), 38–59.

Ying Sng, C., Carter, M. and Stephenson, J.: 2018, A systematic review of the comparative pragmatic differences in conversational skills of individuals with autism, *Autism & Developmental Language Impairments* **3**, 1–24.