Title

Adaptation to Speakers is modulated by working memory updating and theory of mind -- a study investigating humor comprehension

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Loy, Jia
Demberg, Vera

Publication Date
2024

Peer reviewed

Adaptation to speakers is modulated by working memory updating and theory of mind – a study investigating humor comprehension

Jia E. Loy and Vera Demberg
Department of Mathematics and Computer Science
and Department of Language Science and Technology
Saarland University
jia,vera@lst.uni-saarland.de

Abstract

When humans communicate, they typically adapt to their conversational partner in how they speak, and in how they interpret what the conversational partner says. In the area of pragmatic language comprehension, there is so far little work that has studied the individual differences between listeners with respect to adapting to a given speaker. We investigated which individual cognitive factors correlate with listener's ability to associate speakers with humorous utterances. We found that working memory updating (as measured by the Keeping Track Task) was a significant predictor of adaptation to the speaker. These findings are in line with a recent related study (Schuster et al., 2023) which investigated speaker-specific adaptation to the use of uncertainty expressions. We furthermore observe a correlation between speaker adaptation and the Faux Pas Test. This task is used for measuring theory of mind abilities and is believed to specifically tap into intention recognition, an ability which is also very relevant to joke comprehension.

Keywords: figurative language; humour; individual differences

Introduction

In understanding language, comprehenders often have to resolve multiple sources of ambiguity. This may be linguistic-pragmatic in nature, such as ambiguity about a speaker's communicative intentions (Bosco, Bucciarelli, & Bara, 2004), or it may be meta-linguistic, such as ambiguity about a speaker's identity or personality. The two are often inherently connected, for instance, knowledge or beliefs about a speaker's personality may influence pragmatic comprehension such as whether a speaker is being literal or ironic (Pexman & Olineck, 2002; Pexman, 2004).

In addition to making ad-hoc inferences based on stereotypes or beliefs, comprehenders are also known to keep track of speakers' idiosyncratic use of language, and adapt their expectations over time about what the speaker might say (Schuster & Degen, 2019; Regel, Coulson, & Gunter, 2010). Regel et al. (2010) for instance measured participants Event Related Potentials (ERPs) while they read literal and ironic sentences produced by two speakers across two sessions. In the first session, the speaker who was less ironic elicited an increased P600 following ironic statements compared to the speaker who was more ironic. Crucially, in the second session, when both speakers produced an equal number of ironic and non-ironic statements, participants still showed a difference in their ERPs with the two speakers. This suggests that they had adapted to the speakers' communicative styles over time, impacting their subsequent interpretation of irony produced by each speaker.

To date, the majority of research on adaptation to speaker-specific pragmatic language has focused on population-level differences; less is known about how variation at the individual-level may influence such adaptation. There is evidence, however, of a more general link between individual differences and pragmatic language comprehension. In particular, theory of mind (ToM), which is often defined as the ability to infer about the beliefs, intentions, and emotions of others, is known to play a prominent role in various pragmatic phenomena. Bischetti, Ceccato, Lecce, Cavallini, and Bambini (2023) for example observed that older adults with higher ToM skills also showed better understanding of humour, specifically "mental" jokes, which require reasoning about the mental states of joke characters (cf. Samson, 2012; Aykan & Nalçacı, 2018). ToM ability is also known to correlate with other pragmatic phenomena such as the comprehension of irony, implicatures, and indirect requests (Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012; Fairchild & Papafragou, 2021; Trott & Bergen, 2020), highlighting how inter-individual variability in cognitive skills underlies the pragmatic comprehension in communication. A natural follow-up question then is how individual differences may modulate adaptation behaviour to different speakers in pragmatic comprehension.

This question was addressed by Schuster, Mayn, and Demberg (2023), who investigated individual differences in comprehenders' adaptation to ambiguity in speakers' use of uncertainty expressions. In the study, participants were exposed to two speakers who varied in their use of the expressions *might* and *probably* to describe the probability of an uncertain event ("You might / You'll probably get a blue one" in reference to a gumball machine with varying proportions of blue and orange gumballs). In addition, participants completed individual differences tests measuring their working memory (WM), ToM, reasoning ability, and linguistic experience. The study found that participants' tendencies to update their expectations on which expression a speaker would use were correlated with their scores on the Keep Track Task (Yntema, 1963), suggesting that individual differences in WM modulate comprehenders' speaker-specific adaptation abilities.

In the current study, we are also investigating how comprehenders adapt over time to speakers' idiosyncratic commu-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

nicative styles, but we extend previous findings on this question to another aspect of pragmatic language processing – humour comprehension. We focus on individual differences in two aspects of comprehenders' cognitive skills: their working memory and their theory of mind ability. As Schuster et al. (2023) note, their measure of WM could be attributed to two components that the Keep Track Task taps into: WM updating – the ability to hold and modify active representations in memory, and WM storage – the overall capacity for maintaining representations in memory (Ecker, Lewandowsky, Oberauer, & Chee, 2010). WM updating may be relevant as comprehenders have to track and modify speaker representations whilst they receive new speaker-specific input; however, total WM capacity may also be relevant as comprehenders have to maintain a larger number of distinct speaker representations and mappings to communicative styles. To tease apart the potential contribution of the two, we include an additional task, the Operation Span task, which taps into WM capacity but not updating. Additionally, we consider the possibility that adaptation has a social reasoning component: comprehenders who are more inclined to reason about a speaker's mental state may also be more likely to recognise differences in speakers' communicative intentions and pragmatic language use. Although Schuster et al. (2023) did not find a modulating role of ToM on adaptation, it is possible that the ToM task they employed – the Reading Mind in the Eyes Test (RMET; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997) – relies primarily on emotion recognition, an aspect of ToM that is less relevant to adaptation (Quesque & Rossetti, 2020). Thus, in the current study we employ a different ToM task that may be more closely related to adaptation behaviour – the Faux Pas Test (Stone, Baron-Cohen, & Knight, 1998).

Experiment 1

Participants completed a task of humour comprehension and speaker adaptation, followed by an individual differences battery measuring WM updating, WM capacity, and ToM. In the main adaptation task, participants first learnt about two speakers with different communicative styles (literal/humorous) through a series of cartoons. In a subsequent test phase, we assessed how well they had adapted to the two speakers through a task of selecting the correct speaker (obscured) given the dialogue for a new set of cartoons.[1]

Main task: Humour comprehension and speaker adaptation

The main task consisted of two phases: an exposure phase and a test phase. In the exposure phase, participants saw cartoons ($n = 10$) featuring one of two main characters (Peter or Larry) interacting with a secondary character (a friend or family member). The cartoons comprised a prologue, which established the context for the interaction, followed by a dialogue, in which the main character responded to an utterance produced by the secondary character. The responses were taken from the materials of Coulson and Williams (2005), which consists of single-sentence stimuli with final words that elicit a non-joke or a joke interpretation, e.g., "I got a car for my wife, and I thought it was a good choice (non-joke) / trade (joke)". Fig. 1 shows an example of an exposure item. After each cartoon, participants answered a Yes/No comprehension question targeting their understanding of some aspect of the cartoon other than the critical response. Participants saw five exposure cartoons featuring each main character, with each character producing four out of five literal (non-joke) or humorous (joke) responses depending on their communicative style. Character identity and communicative style were counterbalanced across participants.

In the test phase, participants saw similar cartoons ($n = 20$) except that the main character was now obscured (see fig. 1). For each cartoon, participants had to select which of the two speakers they thought produced the response (testing speaker adaptation), and then whether they thought the speaker was "being serious" or "trying to be humorous" (testing speaker intention recognition).

Participants

Three hundred and forty-two native US English speakers were recruited on Prolific for the main speaker adaptation task. They were paid **2.66** GBP for their participation in the main task, which took 15 min. We did not specifically select for neurotypicality or exclude neurodivergent participants. Fourty-four participants were excluded based on our pre-registered criteria of 80% accuracy in the exposure phase. We categorised the remaining participants as learners (further sub-divided in prototypical and reverse learners) or non-learners based on how well they had adapted to the two speakers' different communicative styles. Table 1 shows the breakdown of participants by category. A subset of prototypical learners and non-learners were then recruited for the individual differences battery follow-up ($\sim$ 2 weeks later). We decided to only collect individual difference data for this subset of participants due to the very skewed distribution of participants, which required recruiting a large number of participants per experiment in order to obtain a sufficient number of participants in each category. The experimental costs per participant for the individual difference battery amount to 10.50 GBP per participant for 60 min, totalling 913 GBP plus fees and bonuses for the subset of 87 participants; a bonus was paid if a participant took substantially longer than 60 min, to achieve a rate of 10.50 GBP/h[2].

Table 1: Breakdown of participant categorisation

Experiment	Proto learner	Non-learner	Reverse learner
Exp 1	168	64	66
Exp 2	41	268	19

[1] Pre-registration details available at https://osf.io/jcgdp

[2] Running the full set of participants would have cost at least 3,591 GBP plus fees, exceeding the budget for this study.

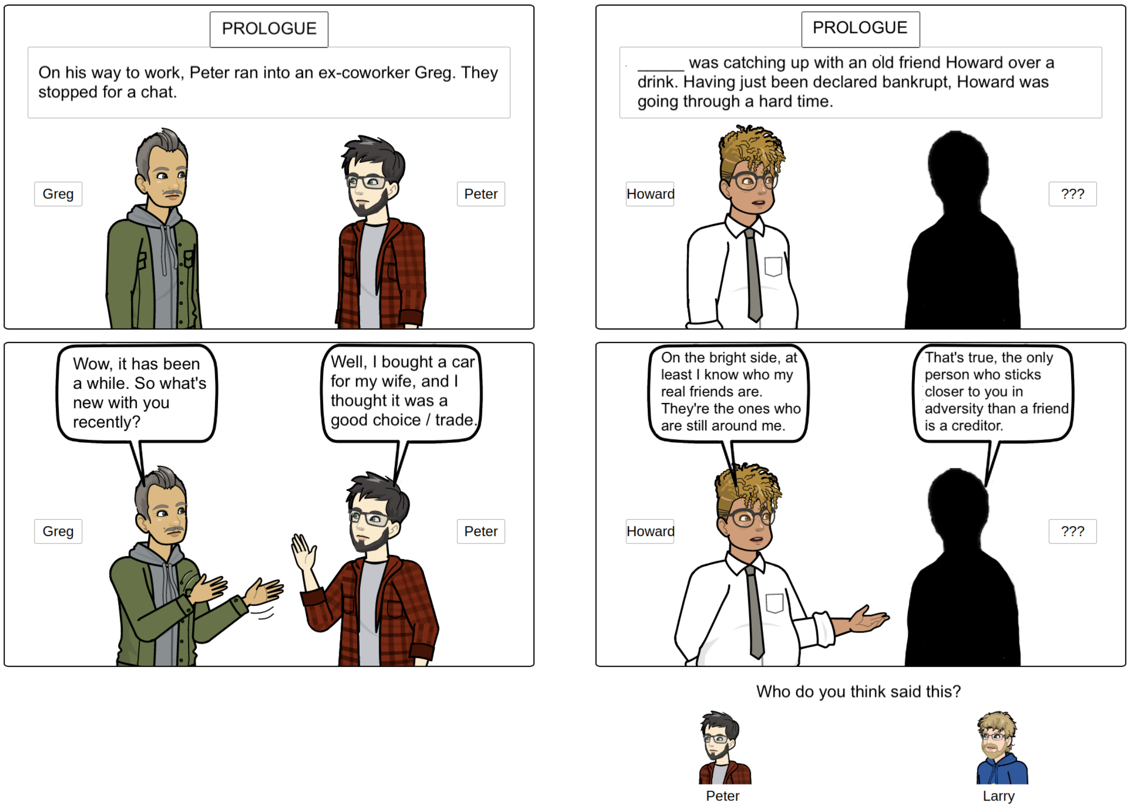

Figure 1: Example of a cartoon shown in the exposure phase (left) and the test phase (right).

Participant categorisation To categorise participants, we first calculated for each participant a speaker adaptation accuracy score based on whether they selected on each test trial the correct speaker (literal/humorous) given their interpretation of the speaker's intentions (being serious/humorous). For instance, a trial on which the participant chose the humorous speaker and also indicated that they thought the speaker was trying to be humorous was scored as correct. An accuracy score was calculated based on the proportion of test trials on which the participant was correct.

To quantify the degree of adaptation, we simulated participant distributions using binomial sampling ($n = 5000$) for three categories of participants. A prototypical learner selected the correct speaker for a given communicative style with the same input probability as in the exposure phase (0.8); a prototypical non-learner selected speakers with a random probability (0.5). An additional category of learner with a selection accuracy of 0.2 was simulated to capture participants who appeared to reverse the mapping of speaker to communicative style, representing participants who adapted to the different communicative styles, but confused the speaker identities.

We then calculated the odds of a participant with a given speaker adaptation score of being a learner (either prototypical or reverse) or a non-learner, followed by the log odds ratio of the two odds. This provides a measure of how much more a given participant's pattern of response resembles that of a learner than non-learner: more positive values indicate greater evidence for being a learner (i.e. having adapted to the two speakers having different communicative styles), more negative values indicate greater evidence for being a non-learner, and 0 indicates equal likelihood of being either.

We defined three categories of participants: prototypical learners ($n = 168$) had a positive log odds ratio score and a speaker selection accuracy of ≥ 0.7; reverse learners ($n = 66$) had a positive log odds ratio score and a speaker selection accuracy of ≤ 0.3; non-learners ($n = 64$) had a negative log odds ratio score and a speaker selection accuracy of between 0.3 and 0.7. From this, we re-invited a subset of representative learners (those with the highest log odds ratio scores; $n = 44$) and non-learners (those with the lowest log odds ratio scores; $n = 43$) for the individual differences follow-up. We did not re-invite reverse learners, and hypothesize that these might be learners who confused the two characters with each other. While we think it would be interesting to investigate this hypothesis and find out what factors are predictive of confusing the characters, it does not directly contribute to the core questions of the present study.

Individual Differences battery

Participants completed three tasks in the order: Operation Span (OSpan; a test of WMC), Faux Pas Test (a test of ToM), and Keep Track Task (a test of memory updating).

Operation Span task We measured WM capacity using an automated version of the OSpan (Turner & Engle, 1989). Participants judged the validity of mathematical equations (e.g., $(1+3)*3=9$), and after each equation were shown a letter to memorise. Equation–letter pairs were shown in set sizes from 3–7, with three sets of each size. After each set, participants had to recall the letters in the order of presentation. A WM capacity score for each participants was calculated using the partial-credit unit (PCU) procedure (Conway et al., 2005). Higher scores reflect better WM capacity. Performance in the OSpan is theorised to be linked to participants' WM processing and storage capacity (Duff & Logie, 2001), but crucially not their WM updating ability.

Faux Pas Test We measured ToM using an automated version of the Faux Pas Test (Stone et al., 1998), a test that involves recognising when social transgressions have occurred. Participants read short stories (10 faux pas; 10 control), and answered a series of questions on whether and why a character had committed a faux pas. A faux pas story might, for instance, describe a scenario where a character criticises her friend's curtains, not knowing that they had been newly bought by her friend. The task calls upon the ability to correctly perceive the emotions and intentions of others – a crucial component of ToM (Quesque & Rossetti, 2020). Participants' performance in the Faux Pas Test is known to correlate with pragmatic language comprehension, in particular the interpretation of sarcasm (Zhu & Filik, 2023). A total score for each participant was calculated using the scoring scheme developed by Zhu and Filik (2023), which awards one point per correct answer on faux pas items. Scores ranged from 19–50, with higher scores reflecting better ToM ability.

Keep Track Task We measured WM updating ability using an automated version of the Keep Track Task (Yntema, 1963), a test that taps into the capacity to keep track of continuously changing information. On each trial, participants were presented with words one at a time from six possible categories, and told to keep track of the last word from two to four categories. The task thus requires participants to maintain and modify representations in their WM with the most recent update. A score for each participant was calculated by awarding one point for each final update per category correctly recalled (Schuster et al., 2023). Scores ranged from 16 to 35, with higher scores reflecting better WM updating ability.

Results

We excluded data from two participants who failed to meet the pre-registered minimum final score of 80% on the OSpan. Thus, the final dataset consisted of 85 participants (43 learners; 42 non-learners). Table 2 shows the mean scores for learners and non-learners in the three individual differences tasks. Our analysis focused on two outcomes of interest: speaker adaptation (how well participants adapted to speakers' communicative styles), and speaker intention recognition (how accurately participants recognised jokes and non-jokes). We hypothesised that participants' WM updating ability would correlate with their speaker adaptation scores, while their ToM ability would correlate with their speaker intention recognition accuracy.

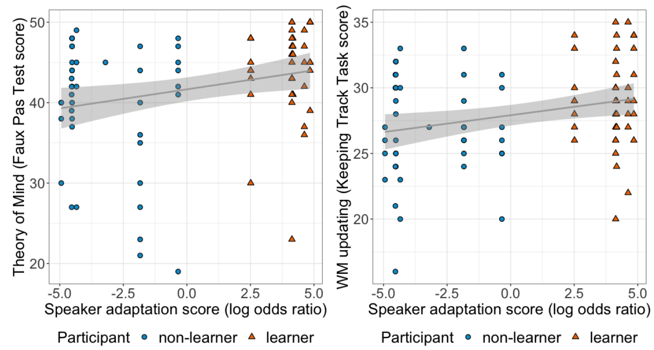

Figure 2: Relationship between participants' speaker adaptation scores and their Faux Pas Test scores (left) and Keep Track Task scores (right).

Table 2: Mean scores in the three individual differences tasks in Experiments 1 and 2 (standard errors in parenthesis)

Experiment	Ppt category	OSpan	Faux Pas	Keep Track
Exp 1	learner	79.0 (18.5)	44.0 (5.3)	29.1 (3.4)
	non-learner	83.0 (11.1)	39.6 (7.9)	27.0 (3.9)
Exp 2	learner	78.7 (18.9)	45.3 (2.5)	28.8 (3.5)
	non-learner	83.2 (11.2)	43.3 (6.6)	28.6 (3.2)

Speaker adaptation We analysed the data using linear regression with participants' adaptation scores as the dependent variable and the three individual differences measures as predictors. Individual differences scores were scaled and centred by converting to z-scores. For the dependent measure, we calculated for each participant who took part in the battery a new log-odds ratio score as a measure of speaker adaptation. We simulated distributions for prototypical learners and non-learners using binomial sampling ($n = 5000$) with speaker selection accuracy probabilities of 0.8 and 0.5 respectively. Following the same procedure as in the participant categorisation, we then computed the log-odds ratio of the odds of each participant being a learner and a non-learner. More positive values indicate stronger evidence of having adapted to the two speakers' different communicative styles.

The model showed an effect of ToM, such that higher scores on the Faux Pas Test were associated with higher speaker adaptation scores, $\beta = 0.91$, $SE = 0.40$, $p = .03$; and of WM updating, such that higher scores on the Keep Track Task were associated with higher speaker adaptation scores, $\beta = 1.06$, $SE = 0.41$, $p = .01$. These relationships are illustrated in Figure 2. We found no significant effect of WM capacity on adaptation ($\beta = -0.53$, $SE = 0.41$, $p = .2$).

Speaker intention recognition For speaker intention recognition, we analysed the data using logistic regression with speaker intention recognition accuracy as the dependent variable and the three individual differences measures (z-scores) as predictors. Recognition accuracy was a binary outcome based on whether participants correctly identified the speaker as "being serious" or "trying to be humorous" for each test trial cartoon. The model showed no effect of any of the three individual differences measures (all $p > .7$).

Discussion

Experiment 1 revealed several findings of interest. Firstly, comprehenders appeared able to resolve speaker ambiguity given the speaker's communicative style. This is evidenced by the large proportion of participants categorised as "learners", and is particularly notable given the minimal amount of training input during the exposure phase (5 trials per speaker). Moreover, our results demonstrate the nature of the individual variability underlying adaptation behaviour. Here, we found that comprehenders' working memory, in particular WM updating, contributed to their adaptation ability. Thus, we replicated Schuster et al.'s (2023) finding that memory limitations play a role in adaptation behaviour. Additionally, we found a correlation between comprehenders' theory of mind and their adaptation ability. This result is inconsistent with Schuster et al.'s (2023) findings, but may reflect the social component of our adaptation task. We come back to these two points in detail in the General Discussion.

In the current experiment, participants encountered ambiguity that was meta-linguistic in nature in the form of an unknown speaker. Such a paradigm is limited, however, in that it is relatively passive since participants merely have to map utterances to speakers. Moreover, the task of identifying an unknown speaker given an utterance is somewhat artificial, as one's interlocutor is typically known in real-life communicative situations. On the other hand, comprehenders frequently encounter linguistic ambiguity in the form of ambiguous utterances or upcoming speech. Such communicative situations are known to influence processing expectations: for instance, listeners interpret anomalous sentences differently depending on the speaker's profile (Brehm, Jackson, & Miller, 2019; Gibson et al., 2017), or they may predict upcoming expressions given their knowledge about the speaker (Trainin & Shetreet, 2023; Hadley, Fisher, & Pickering, 2020). Thus, adaptation in this case may go beyond passive comprehension to active anticipation of what a particular speaker might say. A question of interest then is whether the same individual differences modulate adaptation behaviour to linguistic ambiguity in the form of the speaker's utterance. In Experiment 2, we turn our attention to this question.

Experiment 2

Experiment 2 was similar to Experiment 1 except that the main character in the test phase was no longer obscured; instead, the final word of the speaker's utterance was replaced by a blank, and participants had to complete the utterance for the speaker (e.g., "the only person who sticks closer to you in adversity than a friend is a ____"). Thus, the utterance was potentially ambiguous between a joke and a non-joke depending on what participants thought the speaker might say. As in Experiment 1, participants saw 10 cartoons (5 per literal/humorous speaker) in the exposure phase, and 20 cartoons (10 per speaker) in the test phase.

To evaluate participants' responses, each completion in the test phase was rated on a scale of 1–10 for funniness (1=not funny at all; 10=very funny), calibrated to a rating of 2 and 9 for the non-joke and joke endings from the original Coulson and Williams (2005) stimuli set respectively. The responses were rated by a research assistant who was a native US speaker of English and familiar with the experiment materials but blind to condition (literal/humorous speaker) of each response.

The individual differences battery that followed the main task was the same as that used in Experiment 1.

Participants

Three hundred and fifty-six native US English speakers were recruited on Prolific for the main adaptation task. They were paid 3.11 GBP for the task which took 18 min. Of these, 28 were excluded for low accuracy (<80%) in the exposure phase, and the remaining were classified as prototypical learners, reverse learners, and non-learners as in Experiment 1. Table 1 shows the breakdown of participants by category.

Participant categorisation Participants' ratings on each test trial were first binarised (1–5 vs. 6–10) to classify each completion as a non-joke or a joke. A speaker adaptation accuracy score for each participant was then calculated based on the proportion of test trials on which their response (non-joke/joke) aligned with the given speaker's communicative style (literal/humorous). The subsequent classification of participants employed the same procedure as Experiment 1. A subset of prototypical learners ($n = 25$) and non-learners ($n = 38$) were recruited for the individual differences follow-up ($\sim$ 3 months later) (payment was again 10.50 GPB for the 60 min ID battery).

Results

We excluded data from seven participants who failed to meet the minimum final score of 80% on the OSpan. Thus, the final dataset consisted of 56 participants (22 learners; 34 non-learners). Table 2 shows the mean scores for learners and non-learners in the three individual differences tasks. The main question of interest here was whether and how participants' adaptation scores would be modulated by their individual differences in WM function or ToM ability.

Speaker adaptation We analysed the data following a similar procedure to Experiment 1 by simulating distributions for learners and non-learners via binomial sampling with speaker selection accuracy probabilities of 0.8 and 0.5 respectively. We modelled adaptation using linear regression with the dependent variable of speaker adaptation score calculated by computing the log-odds ratio of the odds of each participant being a learner and a non-learner, and regressed onto the three individual differences measures as predictors. Although the numerical direction of all three effects was the same as Experiment 1 (see table 2), none of the three measures were significant in the model (all $p > .2$).

General Discussion

In the current set of studies, we investigate how adaptation to speaker-specific communicative styles is modulated by individual differences in working memory and theory of mind. We found that comprehenders with higher WM updating and higher ToM ability also showed a greater adaptation effect to the two speaker's communicative styles (Experiment 1). However, these individual differences no longer appear to be relevant when the adaptation task involves an additional step of formulating an utterance in the speaker's communicative style (Experiment 2).

Our results from Experiment 1 are consistent with earlier work that comprehenders adapt over time to speakers' idiosyncratic communicative styles (Regel et al., 2010; Schuster et al., 2023). More importantly, we replicate Schuster et al.'s finding that individuals vary considerably in their adaptation tendencies, and that this variation is modulated by their WM function. The fact that we found an effect of WM updating but not WM capacity indicates that it is specifically the ability to modify speaker representations in response to new input that is relevant to speaker-specific adaptation. Thus, our results add to the literature on speaker-specific adaptation by demonstrating systematic variation at the individual-level, and offer new insight on the role of memory limitations as a cognitive mechanism underlying the adaptation behaviour.

While our findings with respect to WM updating align with Schuster et al.'s, our ToM results differ as Schuster et al. (2023) found no relationship between ToM and adaptation to uncertainty expressions. We speculate that this may reflect differences in the nature of the two ToM tasks: notably, the RMET used by Schuster et al. (2023) is theorised to tap into the emotion-recognition component of ToM, whereas the Faux Pas Test is theorised to tap into belief- and intention-recognition (Quesque & Rossetti, 2020). This may suggest that some aspects of ToM, for instance the tendency to reason about others' mental states, may be more relevant in speaker-adaptation than others. This also aligns with recent theoretical accounts that ToM is not a purely monolithic ability, and different ToM tasks likely tap into different sub-components of the construct (Navarro, 2022; Schaafsma, Pfaff, Spunt, & Adolphs, 2015). Another explanation however may be the pragmatic nature of our adaptation task, which requires comprehenders to understand speakers' use of humour as a prerequisite to mapping communicative style to speaker. Thus, theory of mind, which is known to play a prominent role in pragmatic communication, may be more relevant in adaptation to speakers' use of humour than their use of uncertainty expressions. Future work could provide a more holistic picture of the role of ToM by investigating adaptation to pragmatic and non-pragmatic phenomena. A promising direction for future work also includes comparing neurotypical with neurodiverse populations, specifically speakers with autism spectrum disorder.

Our results from Experiment 2 fail to show evidence for working memory or theory of mind modulating adaptation in a context that involves active production. Here, we speculate that differences in task difficulty due to the linguistic processes called upon in each experiment may have contributed to the lack of finding a significant effect. One notable result that emerges from a comparison of the two experiments is the breakdown of participants across categories. Specifically, Experiment 1 saw a large proportion of learners to non-learners, whereas Experiment 2 found the opposite. Given the exposure phase in both experiments was identical, this difference in numbers likely reflects differences in task difficulty in the test phase of each experiment. In particular, Experiment 2 requires participants to go beyond simply mapping speakers to communicative styles, to actively formulating utterances for a speaker. This is arguably more challenging as it involves an additional step of production after speaker representations have been established in memory. Thus, one likely possibility is that participants may have recognised the difference between the two speakers' communicative styles, but were unable to go the additional step of formulating jokes in the speaker's style. It is possible that cognitive skills such as working memory or theory of mind may, while being theoretically important to speaker adaptation, have a lesser effect on the produced responses than individual differences in properties such as language ability and creativity, which are crucial for coming up with a joke. This would be a useful avenue for future studies to investigate.

Another surprising finding in our study was the relatively high prevalence of participants that we classified as "reverse learners" – they did show adaptation effects but possibly confused the two characters. Future work could also investigate individual factors that lead to a higher susceptibility to confusing characters.

Our study also has some limitations: due to the high cost of collecting individual differnces data, we decided to only collect it for clear subgroups of non-learners and prototypical learners, and did not collect data from people who showed a less clear pattern or the reverse learners. This way, our experiment potentially became more powerful in terms of showing the presence of the effect, but it may overestimate the size of the effect with respect to a more representative sample of the whole population.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 programme (ERC Starting Grant "Individualized Interaction in Discourse", grant agreement No. 948878).

References

Aykan, S., & Nalçacı, E. (2018). Assessing theory of mind by humor: The humor comprehension and appreciation test (ToM-HCAT). *Frontiers in psychology*, *9*, 1470.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child psychology and Psychiatry*, *38*(7), 813–822.

Bischetti, L., Ceccato, I., Lecce, S., Cavallini, E., & Bambini, V. (2023). Pragmatics and theory of mind in older adults' humor comprehension. *Current Psychology*, *42*(19), 16191–16207.

Bosco, F. M., Bucciarelli, M., & Bara, B. G. (2004). The fundamental context categories in understanding communicative intention. *Journal of Pragmatics*, *36*(3), 467–488.

Brehm, L., Jackson, C. N., & Miller, K. L. (2019). Speaker-specific processing of anomalous utterances. *Quarterly Journal of Experimental Psychology*, *72*(4), 764–778.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*, 769–786.

Coulson, S., & Williams, R. F. (2005). Hemispheric asymmetries and joke comprehension. *Neuropsychologia*, *43*(1), 128–141.

Duff, S. C., & Logie, R. H. (2001). Processing and storage in working memory span. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 31–48.

Ecker, U. K., Lewandowsky, S., Oberauer, K., & Chee, A. E. (2010). The components of working memory updating: an experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 170.

Fairchild, S., & Papafragou, A. (2021). The role of executive function and theory of mind in pragmatic computations. *Cognitive Science*, *45*(2), e12938.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological science*, *28*(6), 703–712.

Hadley, L. V., Fisher, N. K., & Pickering, M. J. (2020). Listeners are better at predicting speakers similar to themselves. *Acta psychologica*, *208*, 103094.

Navarro, E. (2022). What is theory of mind? A psychometric study of theory of mind and intelligence. *Cognitive Psychology*, *136*, 101495.

Pexman, P. M. (2004). Social factors in the interpretation of verbal irony: The roles of speaker and listener characteristics. In *Figurative language comprehension* (pp. 209–232). Routledge.

Pexman, P. M., & Olineck, K. M. (2002). Understanding irony: How do stereotypes cue speaker intent? *Journal of language and social psychology*, *21*(3), 245–274.

Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, *15*(2), 384–396.

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain research*, *1311*, 121–135.

Samson, A. C. (2012). The influence of empathizing and systemizing on humor processing: Theory of mind and humor. *Humor*, *25*(1), 75–98.

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in cognitive sciences*, *19*(2), 65–72.

Schuster, S., & Degen, J. (2019). Speaker-specific adaptation to variable use of uncertainty expressions. In *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 2769–2775).

Schuster, S., Mayn, A., & Demberg, V. (2023). Working memory updating modulates adaptation to speaker-specific use of uncertainty expressions. In *Proceedings of the 45th annual meeting of the cognitive science society* (pp. 1213–1220).

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of cognitive neuroscience*, *10*(5), 640–656.

Trainin, N., & Shetreet, E. (2023). Learning speaker-specific linguistic 'style' is mediated by deviance from common language use. *Language and Cognition*, 1–23.

Trott, S., & Bergen, B. (2020). When do comprehenders mentalize for pragmatic inference? *Discourse Processes*, *57*(10), 900–920.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of memory and language*, *28*(2), 127–154.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, *5*(1), 7–17.

Zhu, N., & Filik, R. (2023). Individual differences in sarcasm interpretation and use: Evidence from the UK and China. *Journal of experimental psychology: learning, memory, and cognition*, *49*(3), 445.