Roles guide rapid inferences about agent knowledge and behavior

Aaron Baker, Yarrow Dunham, Julian Jara-Ettinger
Department of Psychology
Yale University

Abstract

The ability to predict and understand other people's actions is critical for real-world social behavior. Here we hypothesized that representations of social roles (e.g., cashier, mechanic, doctor) enable people to build rapid expectations about what others know and how they might act. Using a self-paced reading paradigm, we show that role representations support real time expectations about how other people might act (Study 1) and the knowledge they might possess (Study 2). Moreover, people reported more surprisal when the events deviated from role expectations, and they were more likely to misremember what happened. Our results suggest that roles are a powerful route for social understanding that has been previously understudied in social cognition.

Keywords: action understanding, roles, social cognition

Introduction

Imagine that you are on vacation and, after a day of sightseeing, you come back to your hotel. While waiting in the lobby, someone comes up to you and tells you "Excuse me, I noticed you don't have any clean towels left in your room". If you turned to look at them and saw that they are in uniform and wearing a "housekeeper" badge, you might naturally confirm that they're right and ask if they can put some new ones there. But what if the person were instead another guest?

This example shows that people have rich expectations about what others might know (e.g., whether you have clean towels in your room), how they might behave (e.g., approaching you to tell you about the lack of towels), and what goals they might have in the future (e.g., whether they'd be willing to put fresh towels in your room). This capacity to represent other people's minds—known as a Theory of Mind (Wellman, 2014; Gopnik et al., 1997)—is a hallmark of human social intelligence, allowing us to quickly understand and predict the behavior of others (Ho et al., 2022; Baker et al., 2017).

A critical challenge towards using Theory of Mind, however, is that other people's mental states are unobservable and must be inferred based on people's observable actions. Research into how we accomplish this has found that people can rapidly infer other people's beliefs and desires through an expectation that agents are rational (Jern et al., 2017; Gergely & Csibra, 2003; Jara-Ettinger et al., 2016).

However, as our opening example shows, not all mental state attributions are guided by a process of rational inference from observable action. Here, the two hypothetical agents (the housekeeper and the guest) behaved in identical ways, but we nonetheless had different reactions regarding their mental states.

Intuitively, these inferences are not guided by how they behave, but by our ability to detect whether they are in a *role*.

Roles are pervasive across social life. Consider how, in the span of a single day, you interact with many agents acting under roles: bus drivers, bosses, police officers, baristas, and receptionists, to name only a few. Roles are often associated with predictable beliefs, desires, and goals. Therefore, the capacity to detect when an agent is occupying a role might be a powerful mechanism that people use to rapidly ascribe mental states and predict behavior, even in the absence of any observable previous actions.

While much work has emphasized the importance of role-based structural relations for social reasoning (e.g., Sunstein 1996; Davis et al. 2022), little is known about whether the detection of roles supports rapid—possibly real time—expectations about other people's knowledge and behavior. That is the goal of our paper. To test this, we employed a self-paced reading paradigm in which participants read short passages in a moving window procedure (Just et al., 1982). By strategically changing words in the passages, we can measure differences in people's real-time processing and comprehension of the passage content (Ehrlich & Rayner, 1981; Frank, 2013). Our work is inspired by related work using this approach to reveal automatic Theory of Mind inferences in everyday conversation (Rubio-Fernández et al., 2019). Across two studies, we manipulate the association between a role and the corresponding behavior (Study 1) or knowledge (Study 2). These manipulations reveal how role representations affect people's real-time processing and comprehension of agent behavior.

Study 1

Study 1 first tests whether roles help generate rapid expectations about future behavior. Participants completed a self-paced reading task where we manipulated whether an agent occupied a role or not, followed by an action that was either role consistent or not. All aspects of the study were preregistered unless explicitly noted[1].

Participants

300 participants were recruited on Prolific ($M_{age} = 42.37$, $SD_{age} = 13.69$; 56% female, 42% male, 2% other) and tested using an open-source software for reading time experiments

[1]Materials and pre-registrations for both studies are available at: https://osf.io/xzquk/

Table 1: **Study 1 Conditions.** Each row represents a condition and the columns show variations for each story.

Condition	Convenience Store	Auto Shop	Cafe
Role consistent	cashier takes money	mechanic takes car keys	server takes empty plate
Role inconsistent	cashier takes gum	mechanic takes cell phone	server takes car keys
Non-role	friend takes money	customer takes car keys	customer takes empty plate

Table 2: **Study 1 Passages.** Left shows story set up and right shows test passage. Dashes indicate region boundaries.

Setup	Test Sentence (Split by region)
You are in a convenience store with a friend and you both walk up to the cashier. You put a pack of gum and some money on the counter.	[Your friend / The cashier] – looks down at – the counter, takes – [the gum / the money] – and says – 'Thank you.'
You are at the auto shop and you finish filling out paperwork with another customer and the mechanic. You put your car keys and your cell phone on the counter.	The [customer / mechanic] – looks down at – the counter, takes – [your cell phone / your car keys] – and says – 'Thank you.'
You are at a cafe sitting at a table, surrounded by some other customers and servers. You finish eating a muffin and put the empty plate down next to your car keys.	A [customer / server] – walks up to – your table, takes – [the empty plate / your car keys] – and says – 'Thank you.'

(Zehr & Schwarz, 2018). 22 additional participants were recruited but not included in the study for failing attention checks (n=11) or self-reporting as non-native English speakers (n=11).

Stimuli

Stimuli consisted of three short passages each describing a social interaction happening in locations where roles are common (convenience store, auto shop, and cafe). Each story was created so that it could be modified through minimal wording changes, according to three conditions shown in Table 1–2. In the *role consistent* condition, an agent in a role (cashier, mechanic, server) performed a role consistent behavior. In the *role inconsistent* version, the role agent performed an inconsistent behavior. To ensure differences were due to the mismatch between role and behavior (rather than some behaviors being intrinsically surprising), the *non-role* condition used the same behaviors from the role inconsistent version, but replaced the agent with someone who does not occupy that role. Each passage was broken up into 17 small regions, which were incrementally revealed to participants through a self-paced reading procedure.

We further included three distractor passages (available in OSF repository) designed to prevent participants from picking up on the common syntactic structure across test passages, which participants might then use to skim through. Distractor passages were similar in length to the test passages and also described a social scene, but did not follow the same structure or topics. For example, one distractor was as follows: "You are driving to the mall and you hit a pot hole. You notice your tire going flat so you pull up to a gas station. A woman comes up and offers you her phone to call for help."

Procedure

Participants were told that they would be reading some short passages and answering a few questions. Participants next completed one practice passage to familiarize them with the self-guided reading and comprehension check format. Participants then proceeded to the main phase of the task.

In the task, each participant was presented with three test passages, which they completed one at a time. Participants were randomly assigned to a set of story-condition pairings that ensured they saw each story once (convenience store, auto shop, and cafe) and each condition once (role consistent, role inconsistent, and non-role). For example, a participant might have gotten the following pairing: convenience store + role inconsistent, auto shop + non-role, cafe + role consistent. Participants saw the passages in random order, interspersed with distractor passages (such that each participant completed the three test passages and three distractor passages).

Each passage was split into regions that participants could reveal by pressing the space bar. At the beginning of each passage, participants saw a series of underlines corresponding to each region. When the participant pressed the space bar, words in the first region appeared. As the participant continued to press the space bar, the previous region disappeared and the next region appeared, such that only one region was

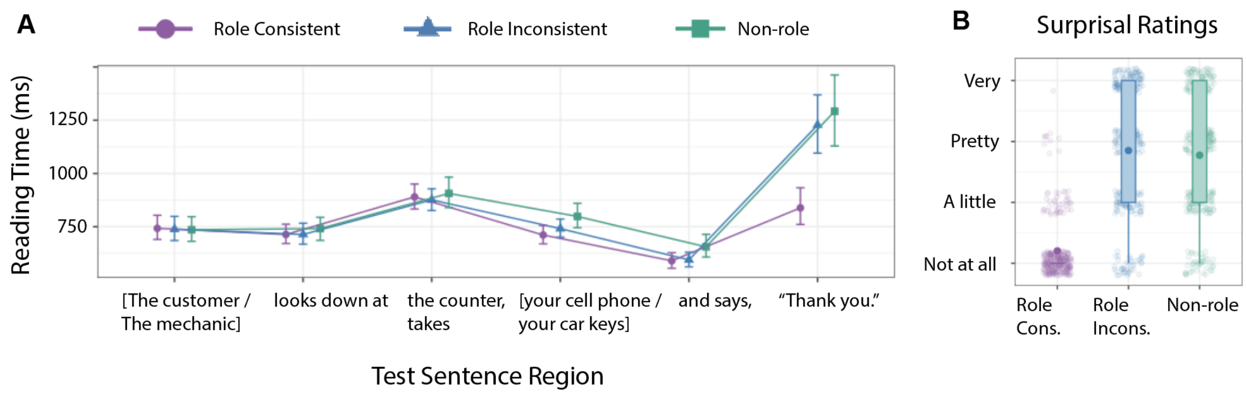

Figure 1: Study 1 results. **A)** Mean reading times and bootstrapped confidence intervals (y axis) for each region of the test sentences (x axis). Results are averaged across all three stories, but x axis shows the auto shop story as an example of the mapping. **B)** Participant explicit surprisal ratings. Box boundaries indicate 1st and 3rd quartiles and dots indicate mean rating.

visible at any time point.

After reading each story, participants were asked two simple objective questions about the story (e.g., "Where did the story take place?"). These questions varied across stories to prevent participants from learning which comprehension checks to expect and hence potentially pay less attention to other features of the story (all questions available on OSF).

After reading all passages, participants answered two memory check questions for each of the three test passages. For each story, participants were reminded of the setup (but not the test sentence) then asked who did the action (e.g., "Who picked up an item from the counter?") with two answer choices (e.g., "The customer" or "The mechanic") and what they did (e.g., "What did they pick up?"; "Your car keys" or "Your cell phone"). The order in which we asked about each story was randomized, but the two questions were always presented in the same order.

After completing the memory check questions participants were asked a direct multiple choice attention check which read "If you are still paying attention please select 'doctor' below." Finally, participants were reminded of how each story ended, and they rated how surprising they found each test passage (4-point likert scale: "Not surprising at all" to "Very surprising") in random order as well.

Results

Figure 1A shows the average reading time, in milliseconds, for each region of the test sentences. Based on our account, we predicted that the role consistent condition would show significantly faster reading times than the role inconsistent and non-role conditions, due to the creation of a real-time behavioral expectation given a role. We pre-registered this prediction to occur in the final region, where the agent says "Thank you" (for all stories). This is because, when an agent takes an unexpected action, participants may expect the confusion to be shortly resolved. By stating thank you, this reveals that this was the agent's final intentional action.

Our pre-registeration indicated that we would exclude participants whose reading time in our target window was two standard deviations above their personal reading speed (mean reading time across all regions of all three test passages). In retrospect this was not appropriate because our prediction is precisely that participants should experience significant slowdowns relative to their average reading speed and hence this method removed the predicted effect, rather than outliers. Thus, here we present results without this pre-registered exclusion criteria, as they are a more accurate reflection of the data and effect sizes. However, all reported effects continue to be significant under our pre-registered analysis.

To test whether participants were significantly faster in the *role consistent* passage, we ran a mixed-effects linear regression with reading time in the target window as the dependent variable, condition (role consistent, role inconsistent, or non-role) as the fixed effect, and random intercepts for participant and story. One response was not recorded due to a software error, therefore this analysis includes 299 observations. As predicted, the role consistent condition revealed significantly faster reading times relative to the role inconsistent passages ($\beta = 391.81$, $p < 0.0001$) and non-role passages ($\beta = 452.96$, $p < 0.0001$).

These effects also emerged within each subject. In an exploratory analysis, we found that most participants showed the fastest reading time (in the target region) in the role consistent condition, despite the story-level effects on reading time. In total, 56% of participants ($CI_{95\%}$=46.82 – 58.19; N=157) had their fastest reading time in the role consistent condition, with the rest of the participants split between the role inconsistent (24% of participants, $CI_{95\%}$=19.40 – 29.10; N=73) and non-role (23% of participants, $CI_{95\%}$=18.06 – 27.76; N=69) passages ($\chi^2 = 49.55$, $p < 0.0001$).

We next analyzed participants' explicit surprisal ratings (Figure 1B) using an ordinal mixed-effects linear regression predicting reported surprise with condition as the fixed effect and random intercepts for participant and story. This anal-

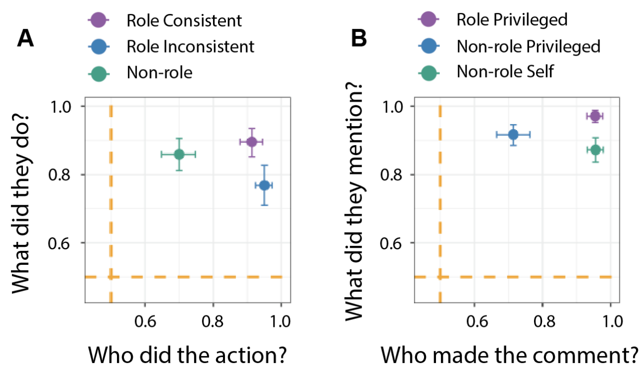

Figure 2: Performance on memory questions for Study 1 (left) and Study 2 (right). Each axis is the proportion of participants who answered correctly for that question split by condition along with 95% confidence intervals. Dashed yellow lines indicate chance performance.

ysis revealed a similar pattern as the implicit reading time analysis. Participants were significantly less likely to be surprised in the role consistent condition compared to the role inconsistent ($\beta = 4.06$, $p < 0.0001$) and non-role ($\beta = 3.77$, $p < 0.0001$) conditions.

We next looked at the results from the memory check questions (Figure 2A). In an exploratory analysis, we examined whether participants would be more likely to correctly remember events in the role consistent condition relative to the other two conditions (inspired by research showing that children produce more memory errors when an agent reports knowing something they shouldn't; Chuey et al. 2023). More specifically, we asked whether participants would misremember which action was taken in the role inconsistent condition (i.e., biased towards role consistent actions) and who took the action in the non-role condition (i.e., biased towards role associated with actions). To test this, we analyzed memory check questions using a mixed-effects logistic regression predicting accuracy as a function of condition, with random intercepts for participant and story. When asked who took the final action, participants in the non-role condition answered with 70% accuracy compared to 91% accuracy in the role consistent condition ($\beta = 1.62$, $p < 0.0001$) and 95% accuracy in the role inconsistent condition ($\beta = 2.28$, $p < 0.0001$). When asked which action the person did, participants in the role inconsistent condition answered with 77% accuracy compared to 90% accuracy in the role consistent condition ($\beta = 1.13$, $p = 0.0005$) and 86% accuracy in the non-role condition ($\beta = 0.74$, $p = 0.01$). Responses for the cafe story were not analyzed for the second memory check question (which action was taken) due to an error in presentation of answer choices.

Discussion

Study 1 shows that role representations shape our expectations about behavior, and this can be seen implicitly through reading times, explicitly through people's self report about surprisal, and through memory errors when asked to reconstruct what happened. Does this extend to associations about mental states? Study 2 explores this question by testing how role representations support expectations about knowledge.

Study 2

Study 2 extends Study 1 to explore whether role representations affect inferences about the knowledge an agent is likely to have. All aspects of the study were pre-registered unless explicitly noted.

Participants

300 participants were recruited on Prolific ($M_{age} = 39.13$, $SD_{age} = 12.96$; 56% female, 42% male, 2% other) and tested using the same experimental software. 12 additional participants were recruited but not included in the study for failing attention checks (n=7) or self-reporting as non-native English speakers (n=5).

Stimuli

Stimuli for Study 2 were structurally similar to Study 1, with three passages describing social interactions in different scenes where people often occupy roles (Table 3–4). The *role privileged* condition describes an agent occupying a role (doctor, housekeeper, guest) revealing private knowledge that is privileged by the role (the ulcer in *your* stomach, the alarm clock in *your* room, the air freshener in *your* car). The *non-role privileged* condition describes an agent outside of a role (patient, guest, customer) making the same comment that reveals knowledge that should be private (*your* stomach, *your* room, *your* car). Finally, to ensure any differences between these two are not due to a mismatch between role and non-role agents generally, the *non-role self* condition describes the same agent outside the role commenting, but instead commenting on some personal knowledge about themselves (*my* stomach, *my* room, *my* car). Each passage was broken up into 18 small regions, which were incrementally revealed to participants through a self-paced procedure. We included the same 3 distractor passages as Study 1.

Procedure

Study 2 followed the same procedure as Study 1, with the only changes being the content of the test passages and their corresponding questions. Like Study 1, participants were assigned to a configuration of story–condition pairings such that they only saw each story once and each condition once. Participants were given instructions and familiarized with the task environment. Participants then read the test passages in random order interspersed with the distractor passages, with all passages followed by two comprehension check questions.

After reading all the passages, participants answered memory check questions following the same format as Study 1. In this study, they were asked who made the comment (e.g., "Who made a comment about the alarm clock being broken?") with two answer choices (e.g., "The guest" or "The

Table 3: **Study 2 Conditions.** Each row represents a condition and the columns show variations for each story.

Condition	Doctor	Hotel	Auto Shop
Role Privileged	doctor, "your ulcer"	housekeeper, "your room"	mechanic, "your car"
Non-role Privileged	patient, "your ulcer"	guest, "your room"	customer, "your car"
Non-role Self	patient, "my ulcer"	guest, "my room"	customer, "my car"

Table 4: **Study 2 Passages.** Left shows story set up and right shows test passage. Dashes indicate region boundaries.

Setup	Test Sentence (Split by region)
You are at a doctor's office in the waiting room where there are donuts and coffee. There is a doctor and another patient drinking coffee along with you.	The [patient / doctor] – looks at you – and says, – "Drinking this coffee – is probably – making the ulcer – in [your / my] stomach – even worse."
You are at a hotel and are waiting in the lobby. You are standing next to a housekeeper and another guest, setting an alarm on your phone for the next morning.	The [guest / housekeeper] – looks at you – and says, – "That might – come in handy because – the alarm clock – in [your / my] room – is broken."
You go to the auto shop to pick up your car. You are at the counter with a mechanic and another customer, and there is a bowl of free air fresheners.	The [customer / mechanic] – looks at you – and says, – "It's a good thing – there are air fresheners, – the one – in [your / my] car – needs replacing."

housekeeper") and what they commented on (e.g., "Whose room did they say the alarm clock was broken in?"; "Your room" or "Their room"). Finally, participants were reminded of the ending they actually read, and they rated how surprising each of the three test passages were (4-point likert scale: "Not surprising at all" to "Very surprising") in random order as well.

Results

Figure 3A shows the average reading time, in milliseconds, for each region of the test sentences. We predicted that the non-role privileged condition would show significantly slower reading times than the role privileged and non-role self conditions. This is because the role would create a real-time expectation of the knowledge that an agent occupying the role might reasonably have (role privileged), but not an agent outside of the role (non-role privileged). However, people should still expect the agent outside of the role to have knowledge about themselves (non-role self). We pre-registered this prediction to occur in the final region for all stories. This is because the final region completes the assertion, and at this point participants see that the agent is not qualifying their use of the knowledge in any way (e.g., by saying "the air freshener in your car *might* need replacing").

To test for reading time differences, we ran a mixed-effects linear regression with reading times as the dependent variable, condition (role privileged, non-role privileged, non-role self) as the fixed effect, and random intercepts for participant and story. As predicted, the non-role privileged condition revealed significantly slower reading times than those reading role privileged passages ($\beta = -161.80$, $p < 0.0001$) and non-role self passages ($\beta = -218.48$, $p < 0.0001$). Using the pre-registered exclusion criteria (same criteria as Study 1) produced the same results, with the difference that the slowdown was no longer significantly higher relative to non-role privileged ($\beta = -33.44$, $p = 0.109$).

In an exploratory within-subject analysis, we found that most participants show the slowest reading time (in the target region) in the non-role privileged condition, despite the story-level effects on reading time. In total, 46% of participants ($CI_{95\%}$=40.27 – 51.68; N=137) had the slowest reading time in the non-role privileged condition, with the rest of the participants split between the role privileged (30% of participants, $CI_{95\%}$=24.16 – 34.56; N=88) and non-role self (24% of participants, $CI_{95\%}$=19.46 – 29.19; N=73) passages ($\chi^2 = 22.56$, $p < 0.0001$).

Participants' explicit surprisal ratings (Fig. 3B) were consistent with the reading times such that participants were significantly more likely to be surprised in the non-role privileged condition compared to the role privileged ($\beta = -1.60$, $p < 0.0001$) and non-role self ($\beta = -2.59$, $p < 0.0001$) conditions (using an ordinal mixed-effects linear regression predicting reported surprise with condition as the fixed effect and random intercepts for participant and story).

For the memory check questions (Figure 2B), we predicted that participants would be more likely to misremember events in the non-role privileged condition relative to the other two conditions. We tested this prediction through a mixed-effects

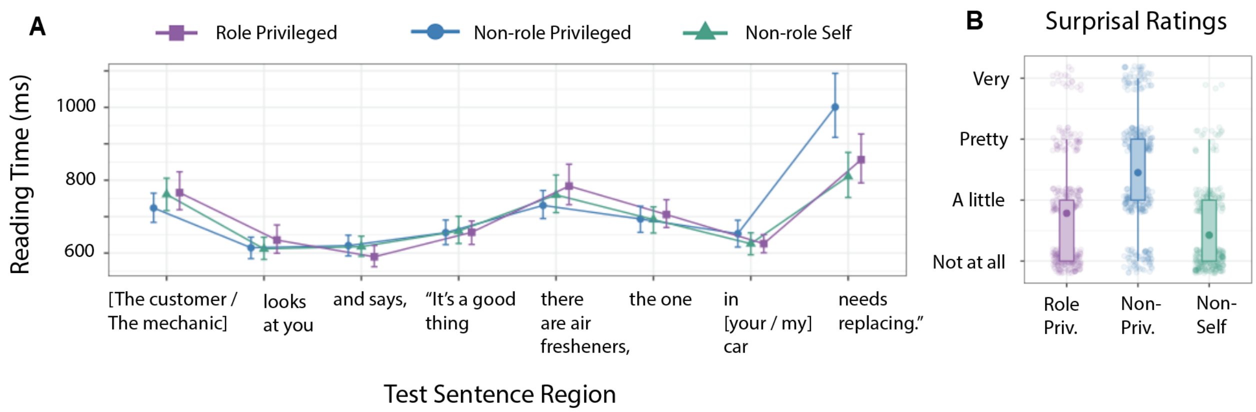

Figure 3: Study 2 results. **A)** Mean reading times and bootstrapped confidence intervals (y axis) for each region of the test sentences (x axis). Results are averaged across all three stories, but x axis shows the auto shop story as an example of the mapping. **B)** Participant explicit surprisal ratings. Box boundaries indicate 1st and 3rd quartiles and dots indicate mean rating.

logistic regression predicting accuracy as a function of condition, with random intercepts for participant and story. When asked who made the comment, participants in the non-role privileged condition answered with 72% accuracy compared to 96% accuracy in the role privileged condition ($\beta = 2.22$, $p < 0.0001$) and 96% accuracy in the non-role self condition ($\beta = 2.24$, $p < 0.0001$). When asked what the person commented on, participants were most accurate in the role privileged condition with 97% accuracy, compared to 92% accuracy in the non-role privileged condition ($\beta = -1.18$, $p = 0.004$) and 87% accuracy in the non-role self condition ($\beta = -1.70$, $p < 0.0001$).

General Discussion

To navigate the social world, people often need to quickly consider how others might act, and what they might know. Our work suggests that this capacity is supported by the everyday recognition of roles. In Study 1, people processed role-consistent actions faster, reported less surprise, and had an easier time remembering what happened. In Study 2, we found the same effects with role-consistent knowledge. Together, these results suggest that people have rich knowledge about roles that they use to build rapid expectations about other people's behavior and mental states.

Our work contributes to research on how people are able to understand each other so quickly. Classical work has often attributed this success to our capacity to infer mental states from observable behavior. This form of mental-state inference emerges early in infancy (Gergely & Csibra, 2003; Onishi & Baillargeon, 2005), and is surprisingly sophisticated in adults (Baker et al., 2017; Jara-Ettinger & Rubio-Fernandez, 2021). However, this form of social cognition suffers the limitations that it requires first observing someone's behavior, and the underlying inference is computationally demanding (Kwisthout & Van Rooij, 2013). Our results suggest that role representations might help solve this problem, allowing us to build quick expectations about other people's actions and mental states.

Our experiments relied on a set of common roles (doctor, housekeeper, mechanic, cashier, and server) that people had behavioral and epistemic expectations about. This opens three questions for future work. First, how complex are our representations of each role? Intuitively, we can make much richer inferences about what people in roles are likely to believe, know, and how they would act than we examined here (e.g., whether a doctor has routine vaccinations). It is possible that our work only tapped into relatively minimal inferences that emerge in straightforward contexts. Second, how broad is our "library" of roles that we use in everyday life? Is it possible that these representations extend to cases that are not technically roles (e.g., we might represent *stranger* as a type of role where its occupant has minimal information about us and engages in small talk). Third, what is the developmental trajectory of this knowledge? That is, how do we come to learn the internal structure of roles? While we do not know the answer to this question, related research has found that children are often focused on so-called *sociodramatic play*—i.e., play where they enact real-world roles—and it is possible that these types of play may support the development of these expectations in adulthood (Jara-Ettinger & Dunham, n.d.).

One limitation of our work is that we do not know the exact nature of the expectations people built from roles. Our work employed a general type of "violation of expectation" paradigm, showing that people were able to maintain a consistent reading speed when roles supported inferences about behavior or knowledge, contrasted by striking slowdowns when this was not the case. Understanding how strong these expectations were, and in which context they are triggered, is a question we hope to pursue in future work.

Altogether, our work highlights the importance of roles for everyday social cognition, and helps explain how people are able to understand and interact with each other in powerful and uniquely human ways.

Acknowledgements

This work was supported by NSF award BCS-2045778.

Thank you to Paula Rubio-Fernandez for support in developing this project, to Aastha Pokharel, Mia Terracino, Tobias Roemer and others for lending us their social intuition, and members of the Social Cognitive Development Lab and Computational Social Cognition Lab for their comments and feedback throughout.

References

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 0064.

Chuey, A., Jara-Ettinger, J., & Gweon, H. (2023). Violation of epistemic expectations: Children monitor what others know and recognize unexpected sources of knowledge.

Davis, I., Dunham, Y., & Jara-Ettinger, J. (2022). Inferring the internal structure of social collectives.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior*, *20*(6), 641–655.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in cognitive science*, *5*(3), 475–494.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292.

Gopnik, A., Meltzoff, A. N., & Bryant, P. (1997). *Words, thoughts, and theories.* Mit Press Cambridge, MA.

Ho, M. K., Saxe, R., & Cushman, F. (2022). Planning with theory of mind. *Trends in Cognitive Sciences*, *26*(11), 959–971.

Jara-Ettinger, J., & Dunham, Y. (n.d.). The institutional stance. *Under review*.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604.

Jara-Ettinger, J., & Rubio-Fernandez, P. (2021). Quantitative mental state attributions in language understanding. *Science advances*, *7*(47), eabj0970.

Jern, A., Lucas, C. G., & Kemp, C. (2017). People learn other people's preferences through inverse decision-making. *Cognition*, *168*, 46–64.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of experimental psychology: General*, *111*(2), 228.

Kwisthout, J., & Van Rooij, I. (2013). Bridging the gap between theory and practice of approximate bayesian inference. *Cognitive Systems Research*, *24*, 2–8.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Rubio-Fernández, P., Mollica, F., Ali, M. O., & Gibson, E. (2019). How do you know that? automatic belief inferences in passing conversation. *Cognition*, *193*, 104011.

Sunstein, C. R. (1996). Social norms and social roles. *Colum. L. Rev.*, *96*, 903.

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Zehr, J., & Schwarz, F. (2018). Penncontroller for internet based experiments (ibex). doi: https://doi.org/10.17605/OSF.IO/MD832