From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Psychology

2021

Emotion Prediction Across Naturalistic Relationship Contexts

Zhao, Zidong

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01mw22v8619

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Emotion Prediction Across Naturalistic Relationship Contexts

Zidong Zhao 赵紫东

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
PSYCHOLOGY

Adviser: Diana I. Tamir

September 2021

© Copyright by Zidong Zhao 赵紫东, 2021. All rights reserved.

Abstract

Emotions change over time, and the temporal dynamics of emotions vary considerably from individual to individual and from group to group. Successful social interactions require people to track this moving target in order to anticipate the thoughts, feelings, and actions of others. In this dissertation, I combine behavioral studies and computational modeling to study how people make emotion predictions that are both accurate and specific to individual partners across three naturalistic relationship contexts. In Chapter 1, I establish that people can indeed make accurate, emotion predictions at a partner-specific level. They do so by relying on specialized knowledge about their partner to account for individual-level emotional idiosyncrasies. Furthermore, I show that how well people make partner-specific emotion predictions is associated with their social success. In Chapter 2, I explore the inferential strategies that support emotion prediction when people encounter new social partners. I show that people can use self-knowledge to effectively approximate other's emotion dynamics. I also show that people are attuned to the utility of their self-knowledge, such that those who are more typical relied on their self-knowledge to a greater extent. In Chapter 3, I investigate how partner-specific emotion prediction can rapidly emerge as new social partners start accumulating specialized knowledge about each other. I demonstrate that people understand the latent structure of how emotion dynamics might vary across individuals in the population. They exploit this structure, and use induction-like mechanisms to quickly individuate their emotion predictions, after learning only a small amount of information about a novel social partner. This work highlights that in order to gain a fuller understanding of the mechanisms underlying predictive social cognition, it is imperative to study it in conjunction with its naturalistic contexts and constraints.

Contents

Abstract **iii**

0 Introduction **1**

1 Accurate Individuated Emotion Prediction and its Social Benefits **8**
- 1.1 Study 1. Emotion Prediction in Close Friends 16
- 1.2 Study 2. Emotion Prediction in New Roommates 30
- 1.3 Discussion 42

2 Rational Egocentric Projection for Accurate Emotion Prediction **50**
- 2.1 Study 3. Rational Egocentrism for Accurate Emotion Prediction 59
- 2.2 Study 4. Replicating Study 3 71
- 2.3 Discussion 78

3 Latent Social Structure Facilitates Partner-specific Emotion Prediction from Sparse Observations **84**
- 3.1 Study 5. Modeling Latent Social Structures of Emotion Dynamics 90
- 3.2 Study 6. Comparing Structured Models Against Baseline 104
- 3.3 Study 7. Dense Sampling of Transition Space 112
- 3.4 Discussion 120

4 Conclusion **127**

References **147**

Appendix A Supplements to Chapter 1 **148**

Appendix B Supplements to Chapter 2 **164**

The following authors contributed to Chapter 1:
Mark A. Thornton and Diana I. Tamir.

Chapter 1 has been previously published:
Zhao, Z., Thornton, M. A., & Tamir, D. I. (2020). Accurate emotion prediction in dyads and groups and its potential social benefits. *Emotion*

The following authors contributed to Chapter 2:
Diana I. Tamir.

Research included in Chapter 2 has been previously presented at the Society for Personality and Social Psychology annual meeting, Virtual Conference (2021)

The following authors contributed to Chapter 3:
Robert D. Hawkins, Diana I. Tamir, and Thomas L. Griffiths.

Listing of figures

1.1 Person-specific and Group-level Accuracies. 22
1.2 Empirical distribution of mismatched accuracy. 23
1.3 Person-specific and Group-level Accuracies from Study 2. 34
1.4 Empirical distribution of mismatched accuracy from Study 2. 35

2.1 Correlation between Egocentrism and Accuracy scores. 66
2.2 Correlation between Egocentrism and Typicality scores. 67
2.3 Correlation between Accuracy and Typicality scores. 69
2.4 Egocentric inference mediated the Typicality-Accuracy link. 71
2.5 Correlation between Egocentrism and Accuracy scores, replication sample. . 74
2.6 Correlation between Egocentrism and Typicality scores, replication sample . 75
2.7 Correlation between Accuracy and Typicality scores, replication sample . . 77
2.8 Egocentrism mediated the Typicality-Accuracy link, replication sample . . . 78

3.1 Model fit at varying number of latent components 98
3.2 Model recovery for the categorical model 101
3.3 Model recovery for the dimensional model 103
3.4 Mean rated and model predict transition probability 110
3.5 Correlations between model-predicted and human-rated differences 111
3.6 Human ratings track dimensional-model predictions 118
3.7 Human-baseline differences track dimension-baseline differences 119

A.1 STudy 1: Percentile scores of correctly-matched accuracies 149
A.2 Study 2: Percentile scores of correctly-matched accuracies 150
A.3 Study 1: Correlation, person-specific accuracy and Friendship quality. . . . 152
A.4 Study 1: Correlation, person-specific accuracy and Friendship duration. . . 153
A.5 Study 1: Correlation, RMET and Friend's perceived closeness 154
A.6 Study 1: Correlation, Group-level accuracy and Loneliness 155
A.7 Study 2: Correlation, Person-specific accuracy and Relationship Duration . 156
A.8 Study 2: Correlation, Group-level accuracy and Days as a group member . 157
A.9 Study 2: Correlation, Group-level accuracy and Social network size 158

A.10 Study 2: Correlation, Group-level accuracy and Support network size . . . 159

B.1 Covariate sharing introduces spurious construct correlations. 166
B.2 Split-bootstrap removes spurious construct correlations. 168
B.3 Split-bootstrap recovers the true construct correlations. 170
B.4 Correlation between Accuracy and Dyadic Similarity 171
B.5 Correlation between Accuracy and Dyadic Similarity, replication sample . . 172
B.6 Dyadic similarity mediated the Typicality-Accuracy link 173
B.7 Similarity and Egocentrism jointly mediated the Typicality-Accuracy link . 174
B.8 Dyadic similarity mediated the Typicality-Accuracy link, replication sample 175
B.9 Similarity and Egocentrism jointly mediated the Typicality-Accuracy link, replication sample 176

To my parents.

致我的父母, 赵俊锋和谢旭.

Acknowledgments

The throes of second dose side effects seem like a perfectly apt place to acknowledge what a strange and wondrous five years it has been. The chills and the muscle pain have somehow given me the respite so that I can have a moment of reflection and accounting to realize how much I have learned, and give thanks to all those who made it possible.

First and foremost, I am deeply indebted to my advisor Diana Tamir, who has been so incredibly generous with her time and mentorship. I cannot overstate the impact Diana has had since she took a chance on me five years ago when I was, in retrospect, laughably green. She has exemplified (though I am still learning) how to be rigorous yet charitable as a researcher. She has shown me how to bridge nebulous, philosophical questions about the mind with well-defined, executable psychology research. She has taught me how to distill complex ideas and make clear expositions in my scholarly communication. Perhaps most importantly, she had faith in me even when I thought none was warranted.

I would like to express my gratitude to my secondary advisor Tom Griffiths, whose work inspired me to thinking about human learning in a way I did not know before. I approached Tom with a half-formed idea and next to no experience with computational approaches, and I was greeted with nothing but patience and knowledge.

I want to thank Alin Coman and Casey Lew-Williams for taking on the additional labor of being on my committee and sharing their insights, even with all that's going on in the world.

I want to give special thanks to Mark Thornton, whose research inspired much of the work in this dissertation, for being the best mentor and collaborator one could hope for, for quite literally knowing everything, and also for sitting on my committee.

I am also grateful to other collaborators who have also graciously taken on the responsibility of mentorship: Robert Hawkins, Elizabeth McDevitt, and Meghan Meyer. Thank you for embodying the kind of scientist I want to be - relentless, meticulous, learned, and yet inventive.

I want to thank my undergraduate advisors for opening up the world of research to me: Sohee Park, Bunmi Olatunji, Sean Polyn, Megan Saylor, and Jo-Anne Bacharowski.

I would like to extend my sincere thanks to members, present and past, of the Princeton

Social Neuroscience Lab: Judith Mildner, who is family in every sense of the word and deserves her own dissertation-length acknowledgement, for being here with me through all of it; Kara Enz, for bringing me along on so many adventures and welcoming me into every single property she owns; Abla Alaoui-Soce, for being my "we get each other" person, it's meant a lot; Lily Tsoi, for being not only a brilliant mentor and colleague, but also the big sister I have always wanted; Shannon Burns, who's fabulous both in science and on YouTube, for showing me that you can truly have the cake and eat it too; Aaron Kurosu, Nicky Jones, Jordan Rubin-McGregor, Elyssa Barrick, Stats Atwood, Dominic Burkart, and Miriam Weaverdyck, among others, for having been an integral part of my support network. A special acknowledgement is also due to Morrison Mildner and Sunny Enz, who are joy.

I'm extremely grateful to my cohort, without whom I wouldn't have survived even a year: Lauren Feldman, for showing me courage and conviction, as well as the importance of being financially responsible; Sammy Floyd, for consistently challenging me to interrogate how I see things, in the best way possible; Lena Skalaban, for having seen me in my true element; Karina Tachihara, for being the glue and creating a community; Madalina Vlasceanu, for including me in some of the most beautiful, important moments of her life; Xiaofang Yang, for that whole cheesecake on my 24th birthday, among other things; Joe Avery, Branden Bio, Nate Cheek, Jordan Starck, and Sarah and Andrew Wilterson.

I want to thank other Princeton psychology friends "who are technically not in the cohort, but it's more complicated than that": Emily Liquin, for being the calming, steady voice of reason; Fred Callaway, before whom I thought "the absent-minded professor" was a trope; and Mira Nencheva, who is not so secretly the smartest of all of us.

I want to thank the Psychology department staff: Keisha Craig, Jill Ray, Beth Porter, David Carter, Sara Lovelace, Tina McCoy, among others.

I want to thank Linda Chang and Jason Chin, among others, for being the reason I go to SPSP.

I must also thank my extended McTyeire crew: Marissa Davis, May Braverman, Jeff Tang, Jenn Li, Subham Agarwal, Brandon Darr, Martha Babbit, Sophia Cramer, Stephen Zaksewicz, Maddie Manaker, Katie Linek, Elizabeth Linton, and Jyotirmoy Mandal, whose continued friendships make me boundlessly grateful and proud. We are all the journeywork of the stars.

I want to thank Austin Mueller for the all the books I am still planning to read.

I want to thank Junyi Chu for being a role model, both in science and in life.

I want to thank Vivian Liang for being my ride or die, no matter what continent.

I want to thank Ina Yang for being a guiding light. I want to thank Jiawei Lin, Zhuoyang Zhan, Xiao Nie, Tuo Huang, and Zhongyu Wang, for the decade-long group chat.

I also cannot finish this journey without thanking the people of Rojo's Roastery: Kevin,

Matt, Gaven, Cloud, Michelle, Claudia, Gabe, Rowan, Aisha, Rosemary, Andre, Josh, Lauren, Elia, CJ, and so many more

Finally, I cannot begin to express my thanks to my parents, for making all this possible through their hard work and sacrifice. I love you.

0

Introduction

Our social world is forever in motion. Be it a gathering of friends or a midnight crowd of strangers at a neighborhood deli , we routinely find ourselves in the presence of other people, and surrounded by a flurry of information and social demands. Even when we are alone, we are not; when it comes to human sociality, out-of-sight is rarely out-of-mind (Mildner & Tamir, 2018). People ruminate over past interactions, wondering if they could have done anything differently, and they daydream about things that have not yet

happened, envisioning and rehearsing important conversations that might take place.

At first glance, our social interactions can be reduced to mere sensations. People see others' expressions and movements, and they hear the sound of others' speeches. However, the constant stream of social sensory inputs can easily overwhelm, so they must be organized in a principled way if people are to think and act in a cogent, adaptive manner. To this end, people intuitively understand that what can be perceived about others through our senses is in fact generated by latent, unobservable causes. People explain observed behaviors in mentalistic terms, attributing other's actions to beliefs, desires, and emotions. These latent, explanatory notions of the mind confer critical, instrumental utility, as they make it possible for people to predict how others might act in the near future (Dennett, 1989; Heider & Simmel, 1944). By forming accurate expectations about others' thoughts, feelings, and behaviors, people might be able to plan their own actions accordingly, and thus have more success in their social interactions and relationships.

While it is well acknowledged that people recognize mental states as both causes and consequences of the external events immediately surrounding them, recent research has also suggested that mental states, particularly emotions, can provide predictive value beyond the present moment. That is, people can capitalize on the temporal dynamics of emotions to make long-range social predictions. Emotions are not independent, segregated events (Barrett, 2017). They do not occur in a vacuum of previous contexts. Rather, a person's emotions at a given moment hinge on how their preceding emotion states unfold (Cunningham, Dunfield, & Stillman, 2013). Importantly, there are regularities in how emotions change over time. Stress and anxiety might be more likely to morph into frustration rather than joy, while hopefulness might more easily transition into content than

despondence. Using an understanding of these emotion dynamics, people can predict others' future emotions based on knowledge of their current states.

Previous research has suggested that people indeed understand the regularities of emotion dynamics at the aggregate level. People can accurately predict the likelihood that an abstract, average target would transition between two emotions (Thornton & Tamir, 2017). This work foundationally demonstrated people's ability to predict emotion transitions, and established emotion predictions as a novel domain of social cognitive skill. While this research provided strong evidence for emotion prediction in its idealized form, it nonetheless endured limitations to its ecological validity. Like much research studying social cognitive phenomena at the aggregate level, participants made predictions not about a realistic social target, but rather an averaged, abstract entity (Ickes, Robertson, Tooke, & Teng, 1986). Studying social prediction with such an artificial target means that researchers lose out on nuanced information that might be critical for how predictions function in the wild.

Two features of naturalistic social cognition are particularly important for recovering the nuances of emotion prediction in everyday life. The first feature is the individual variabilities in emotion dynamics. In their daily social lives, people need to make predictions about specific social partners, rather than the general, abstract target often used in conventional social cognition tasks. While emotion dynamics might follow a general pattern at the population level, social groups and specific individuals vary in how strongly and how frequently their emotion states oscillate (Kuppens, Oravecz, & Tuerlinckx, 2010). This variability makes emotion prediction more challenging, as people likely cannot accurately predict the emotions of every partner by simply applying their general knowledge in a one-size-fits-all manner. Rather, they might need to cognitively resolve individual variabilities, and adapt

their predictions based on the information they have available about their partner.

The amount of partner-specific information one has is, in turn, determined by the second feature of naturalistic social cognition –the developmental stage of people's relationship with each specific partner. People interact with a large number of other individuals throughout their lifetime, and they necessarily maintain a range of different types of relationships. While people might remain superficially affiliated with most social partners, some of these relationships develop into meaningful, lasting connections. Over the course of relationship development, people can have drastically different types and amounts of information about their partner. Differences in information can then place additional constraints on the already challenging task of emotion prediction, and even its outcome (Klein, Loftus, & Burton, 1989). If people meet someone through a mutual friend at a party, they may fully intend to leave a positive impression by anticipating how their new partner think and act. Nonetheless, they have nothing to base their predictions on other than what can be glimpsed on the surface, and might resort to stereotypical beliefs. On the other hand, friends who share a decades-long relationship might have weathered so much together, they can predict exactly how the other would react to something novel simply by relying on episodes of past experiences.

In this dissertation, I attempt to expand our understanding of emotion prediction into its naturalistic contexts. To account for the cognitive challenges presented by the need to make accurate emotion predictions at a partner specific level, I study emotion prediction in spontaneously occurring dyads from a US university community. To account for the information constraints presented by the nature of different interpersonal relationships, I study emotion prediction across three relationship scenarios –established close friends, new

acquaintances with little to no partner-specific knowledge, as well as new acquaintances at the early stage of accumulating partner-specific knowledge.

In Chapter 1, which was based on previously published work (Zhao, Thornton, & Tamir, 2020), I establish that people can account for others' emotional idiosyncrasies to make accurate, partner-specific emotion predictions. Across two studies, I investigate emotion predictions in two different types of dyads. Participants predicted the emotion transitions of a specific partner –either a close friend (Study 1) or a first-year college roommate (Study 2) –as well as an average group member. Results from these studies demonstrate that people can make highly accurate predictions both for specific individuals and specific groups. I show that accurate predictions rely on partner-specific knowledge. New community members were able to make accurate predictions at zero-acquaintance, but accuracy increased rapidly over time as individuals start accruing knowledge. Furthermore, I show preliminary evidence suggesting that accurate emotion prediction is associated with social success in both individual and communal relationships.

In Chapter 2, I investigate the inferential strategies that support emotion prediction when people encounter new social partners. People often do not have enough information about a novel partner to make accurate inferences. Rather, they often resort to an egocentric heuristic, and make predictions about a target by using their own self-knowledge as a proxy. It is debated whether this egocentric heuristic should be considered a form of cognitive bias or a rational strategy for naturalistic emotion predictions. Using an emotion prediction task and personality measures, I assess interpersonal similarity and predictive accuracy between first year college students and their new acquaintance roommate. I show that, when people need to predict an unfamiliar target's emotions, self-knowledge can often

effectively approximate knowledge about others, and thus support accurate predictions. Moreover, participants that were typical of the sample, whose self-knowledge can better approximate information about the target, relied more on self-knowledge in their predictions, and thus achieved higher accuracy. The findings in this chapter suggest that people rationally tune their use of egocentrism based on whether it is likely to pay off, thus demonstrating a rational side to egocentric emotional predictions when its natural decision context is taken into consideration.

In chapter 3, I combine computational modeling and behavioral experiment to explore how partner-specific emotion prediction can rapidly emerge as new social partners start accumulating specialized knowledge about each other. Using probabilistic models, I first formalize two possible latent structures underlying individual variability in emotion dynamics. I demonstrate that models with a dimensional underlying structure outperformed those with a categorical structure in capturing how people in our samples differed in their emotion dynamics. Furthermore, I examine whether people can exploit these structured regularities to rapidly learn partner-specific emotion dynamics. In an inductive learning task, I ask both human participants and the computational models to predict a novel target's unobserved emotion transitions after receiving only a single observation about their emotion dynamics. Across two behavioral studies, I show that people use inductive reasoning to quickly refine their understanding of partner-specific emotion dynamics. Furthermore, people's inductive reasoning conforms to the prescriptions of the dimensional model, suggesting that they can encode and utilize latent statistical regularities to learn partner-specific emotion dynamics both quickly and accurately.

Finally, I discuss how this dissertation contributes to the extant literature by studying

a novel aspect of social cognition in conjunction with the constraints from its naturalistic contexts. I suggest several avenues for how future research can flesh out our understanding of emotion prediction, as well as social cognition at large, through naturalistic disaggregation.

1

Accurate Individuated Emotion Prediction and its Social Benefits

Prediction is central to social cognition. Whether in cooperation or competition, social interactions require people to anticipate others' future thoughts, feelings, and actions, and prepare their own actions accordingly (Tamir & Thornton, 2018). The ability to intuit another person's social future should confer immense social advantage, as it

should aid action planning, and thus facilitate smoother social interactions. Yet predicting the future feelings of any specific individual is a challenging task. Emotion dynamics vary considerably from individual to individual and from group to group. The same rainy weather might lead to lasting feelings of negativity (and damp socks) for a colleague who tends to wallow, but only fleeting annoyance for one who is more optimistic. To truly reap the social benefits afforded by accurate social prediction, people must tailor their predictions to the target they are predicting. To what extent do people tune into these idiosyncrasies to predict how others' emotions evolve over time? In the current study, we investigate the hypotheses that people can accurately predict specific individuals' future emotions, and that accurate social predictions provides real world social benefits.

1.0.1 Predicting Emotion Transitions

Emotions are not isolated events occurring in a vacuum (Barrett, 2017). Instead, a person's state at any given moment depends on how their previous states unfold over time. State transitions can take place through multiple potential pathways. Some take place due to internal causes, as people actively regulate their emotions and shift into a different state (Gross, 2002). Other state transitions occur because new external events take place, and how these new events lead to internal experiences will be conditioned on the previous state (Rinck, Glowalla, & Schneider, 1992). Marginalizing over these different mediating factors, emotion dynamics often follow predictable patterns (Cunningham et al., 2013), so people can use what they know about how other people feel in the moment to predict how they might feel in the future. Someone who is currently feeling anxious is more likely to become frustrated next rather than joyous, whereas someone who is currently feeling hope-

ful is more likely to feel joyous next rather than frustrated. People can learn how emotions evolve over time both by experiencing these regularities in their own emotional lives and by observing these regularities in others. People can then leverage knowledge about these regularities to form accurate predictions about others' future states. Previous research has found that people are indeed highly accurate when they are asked to predict how a *general target* transitions from state to state (Thornton & Tamir, 2017). That is, people's predictions about the average person mirror the ground truth about how emotions transitions occur in the population.

In everyday life, people need to make predictions about specific individuals, rather than a general target. While emotion dynamics follow predictable patterns in general, individuals vary greatly in both how strongly and how frequently their emotional states fluctuate (Kuppens et al., 2010). These individual differences in emotion dynamics are robust and stable (Heller et al., 2015; Davidson, 2004). Thus, in order to make accurate person-specific predictions, it is unlikely that people indiscriminately apply their general understanding of emotion transitions to an individual. A complete model of social prediction must explain how people make predictions about the emotion transitions of specific individuals.

People likely constrain their predictions by drawing on relevant knowledge. Prior research has shown that people make social inferences by drawing on different sources of knowledge depending on the features of the target. For example, people are more likely to draw on knowledge about themselves to make inferences about similar others, and stereotypes for dissimilar targets (Tamir & Mitchell, 2013; Ames, 2004a). People also fine tune their social inferences based on familiarity with target individuals (Welborn & Lieberman,

2014). Friends are more accurate than randomly paired strangers at inferring each other's thoughts and feelings. This is not because they are more similar, but because they have superior individualized knowledge about each other (Stinson & Ickes, 1992). Having specialized knowledge about a target thus increases one's inferential and empathic accuracy.

In the current study, our primary goal is to answer two questions about social predictions in the domain of emotions. First, we test whether people can make accurate predictions about specific individuals, such as a friend or a new roommate. We do so by comparing person-specific predictions against the target's self-reported emotion transitions. In prior work, self-reported emotion transitions correlated strongly with ground truth transition probability (Thornton & Tamir, 2017). This indicates that they are a reasonable choice against which to benchmark others' predictions. Second, we test whether people make accurate predictions by invoking person-specific knowledge, and not simply their understanding of general emotion dynamics, nor knowledge about their own emotion dynamics.

1.0.2 Social Perception and Its Benefits

People's mental states –their thoughts, feelings, beliefs, and emotions –shape how people behave. An angry person is more likely to aggress than a happy person. Thus, being able to perceive others' current states should help perceivers to gain strategic command over their own social surrounding. Indeed, people seem highly adept at inferring others' unobservable inner worlds. This mind-reading feat has been a central fascination in social psychology, and has been broadly studied across various literatures such as empathy (Decety & Jackson, 2004), emotion recognition (Martinez, Falvello, Aviezer, & Todorov, 2016), empathic ac-

curacy (Ickes, 1993), and theory of mind (Gopnik & Wellman, 2012; Gordon, 1986). This literature indicates that people can indeed accurately perceive others' mental states (Ickes, Stinson, Bissonnette, & Garcia, 1990; Zaki & Ochsner, 2011). This coincides with findings from the person perception literature showing that people form accurate impressions of others' personality traits (Funder, 1995).

Social perception research also indicates that the ability to accurately infer others' ongoing mental experiences is associated with positive real-world social outcomes. Children who can more accurately track the beliefs and intentions of multiple agents enjoy greater acceptance from their peers (Banerjee, Watling, & Caputi, 2011). Adolescents who can more accurately infer the specific contents of others' ongoing thoughts and feelings enjoy better general social adjustment and greater peer acceptance (Gleason, Jensen-Campbell, & Ickes, 2009); adults who do so enjoy greater satisfaction in romantic relationships (Sened et al., 2017). Thus, there is a clear link between social perception and social success.

However, our social world requires that people go beyond reactive inferences about others' ongoing or past emotions. Much like in a game of chess, where seeing multiple steps into the future enables strategic planning, gazing into others' emotional futures would allow people to project their own courses of action further in time. People that can anticipate others' emotions before they occur should benefit from an even greater strategic command over their social surroundings. Such predictions are intertwined with perception in that people might be able to foretell others' future emotions using perceivable information about their ongoing ones.

1.0.3 The Benefits of Social Prediction

We suggest that social prediction should be a powerful driver of social success. Naturalistic social interactions require people to make continuous, rapid social inferences as they unfold in real time. Yet time pressure often undermines people's social inferences (Epley, 2004; Gershman, Gerstenberg, Baker, & Cushman, 2016). While people are adept at rapidly forming impressions (Ambady, 2010; Willis & Todorov, 2006) and perceiving emotions (Martinez et al., 2016) from observation, social prediction can serve as an additional mechanism for offloading such pressure. Anticipating others' thoughts and feelings a few steps ahead can prime people for interactions well before they take place, and knowledge of emotion transitions should allow people to do precisely this. That is, one's ability to use the knowledge they have in the moment to make accurate predictions about the future should be an important springboard for social success in at least two ways.

First, accurate social predictions should help people form stronger relations with individual friends. As relationships develop, people have repeated opportunities to interact with each other. To the extent that people can learn their friends' idiosyncratic emotional experiences, and learn how to accurately predict how they might think or feel, they should be able to have higher-quality interactions at each turn, and therefore enjoy more successful relationships. Here we test whether people who can make more accurate prediction about a friend's emotion transitions also enjoy more success in their relationship.

Second, accurate social predictions should help people interact more successfully with their local community. While well-known others and close friends constitute a large part of people's social environment, people's daily social interactions are not restricted to these individuals. People's social success also depends on the success of their interactions within the

broader social milieu. Groups exhibit idiosyncratic norms around emotions; they can vary widely in how members should emotionally respond to different events, how emotions should be displayed, and which emotions they value (Kolb, 2014). For example, American cultures value high-arousal states and low-arousal states similarly, whereas East Asian cultures value low-arousal states more than high-arousal ones (Tsai, 2007). Group members actively maintain their group membership by feeling and displaying the "correct" emotions (Kolb, 2014; Mesquita, Boiger, & De Leersnyder, 2016). In order to make accurate predictions about unfamiliar group members, people must learn, and then take into consideration the emotional norms and dynamics of their communities. If one's group-level understanding is closely tuned to the regularities of their social group, they could accurately predict others' emotion transitions, even at zero acquaintance. This would facilitate smooth initial interactions, and downstream friendship formation. Here we first test whether people can make accurate predictions of emotion transitions for their current social community; we next test whether people who are more accurate about their group's emotion dynamics enjoy more general social success.

Social perception and social prediction likely both contribute to social success. However, the two might contribute for different reasons. Perceiving emotions in the moment requires that people translate between observed expressions and unobservable mental states, a skill that may rely on using either theory-like knowledge (Skerry & Saxe, 2015) or first-person simulation (Waytz & Mitchell, 2011), and depends on both context and culture (Barrett, Mesquita, & Gendron, 2011; Gendron, Roberson, Vyver, & Barrett, 2014). Predicting how emotions change over time requires that people not only perceive other's current emotion, but also track the emotion dynamics of their social milieu, a skill that

might rely on statistical learning (Newport & Aslin, 2004a). Here we use a conventional measure, the Reading the Mind in the Eyes Test (RMET; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997), to test to what extent emotion perception ability might align with emotion prediction ability. We also test whether perception and prediction differentially predict social success.

1.0.4 The Current Studies

In two studies, we examine emotion predictions within the context of a local university community. In both studies, participants predict emotion transitions for three different targets: their study partner, the average student at their university, and themselves. In our primary analyses, we test whether people make accurate person-specific predictions using tailored person-specific knowledge. Next, we test whether people make accurate group-level predictions. Finally, we test whether individual and group-level accuracy are associated with relationship success and general social success, respectively.

In Study 1, we recruited a sample of close friends who had lived in the same community for at least a semester. We aimed to establish that people can make specific and accurate predictions about their longstanding friends and established community groups. These data also provide a first test of whether the accuracy of one's emotion predictions is associated with any metrics of social success. In Study 2, we recruited a sample of newly acquainted roommates. These data thus allow us to examine the speed with which people can learn to generate accurate predictions, as well as the extent to which social benefits of accurate predictions emerge over time. We expected that new dyads would successfully predict their partner and their community, though to a lesser extent than established dyads. Further, we

expected that the relation between accuracy and social success will not yet have emerged at the very early stage of relationship.

1.1 Study 1. Emotion Prediction in Close Friends

1.1.1 Methods

Code and data availability. Data and code from this study have been deposited on the Open Science Framework (https://osf.io/fmkj7/) and are freely available. We report how we determined sample size, all data exclusions, all manipulations, and all measures in the study. All null-hypothesis tests are two-tailed.

Participants. Forty-seven same-sex dyads (N = 94, 72 women and 22 men) who identified each other as close friends were recruited through SONA from the Princeton University undergraduate population ($M_{age} = 19.68, SD = 1.60$; 50% White, 29.8% Asian, 12.8% Black or African American, 5% more than one race, and 2% unreported; 15.8% Hispanic or Latino). A target sample size of 43 dyads was set a priori; a power analysis based on prior research showed that this sample size would provide 95% power to detect the expected accuracy of participants' emotion predictions ($d = 0.39$; Thornton & Tamir, 2017). The current sample exceeded the a priori criterion set by this power analysis to allow for the exclusion of participants who did not complete the study as instructed. Due to technical difficulties, 3 dyads did not complete the Reading the Mind in the Eyes Test. These dyads were excluded from analyses involving the RMET. One participant provided predictions for the group with zero variance. This participant was excluded from all analyses that include group-level predictions. No data exclusion occurred otherwise. All participants in

this and all subsequent studies provided informed consent in a manner approved by the Princeton University Institutional Review Board.

Emotion transition tasks. Dyads arrived at the laboratory together, but were tested separately. Participants first completed the emotion transition task. In this task, participants judged how likely it was that one state (e.g., happiness) would transition into another state (e.g., elation) for a particular target. On each trial, participants saw a target (e.g., self) and pair of emotional states (e.g., happiness → elation). Participants used a continuous scale from 0% to 100% to rate the likelihood that, if the target were currently experiencing the first state, they would next experience the second state. The states used in this task consisted of three independent sets of five emotional states. To select these states, we applied k-medoids clustering to three large sets of transitional probability data from previous research (Thornton & Tamir, 2017). The state at the center of each cluster was chosen. This selection process ensured that the emotional states used in our task spanned the state space, such that results would not be limited to certain sub-regions of the space (Table 1.1). Stimuli consisted of all possible state pairs within each set (See A for details on state selection). Participants completed a total of 75 trials. Trials were randomized within target blocks.

Set 1	Set 2	Set 3
confident	nervous	satisfaction
grouchy	irritable	love
sad	lively	contempt
assertive	bold	disgust
unrestrained	talkative	embarrassment

Table 1.1: Emotion words included in the emotion transitions task.

Each participant completed the emotion transition task three times, providing transi-

tion ratings for three targets: (i) Self ratings: a participant's ratings of their own emotion transitions, (ii) Friend ratings: a participant's ratings of their friend's emotion transitions, and (iii) Group ratings: a participant's ratings of the emotion transitions of the average student at their university. The three targets were presented in blocks, in random order across participants.

The dyadic design allowed us to assess the ground truth of actual Friend and Group transitions, respectively: (i) Friend-Actual: a participant's friend's self-reported ratings of their own emotion transitions, and (ii) Group-Actual: the sample's average self-ratings, calculated separately for each participant to exclude themselves and their friend.

We tested whether people could make accurate person-specific predictions in three ways. First, we calculated bivariate correlations between participants' predictions for their friend (Friend ratings) and the actual transitions provided by their friend (Friend-Actual). These correlation coefficients were then Fisher's z-transformed, and subjected to a one-sample t-test. Following previous research, we benchmarked our labels for accuracy against Cohen's convention for effect size (Jussim, Crawford, & Rubinstein, 2015). We report raw correlations greater than .4 as accurate, those between .25 and .4 as moderately accurate, and those below .25 as inaccurate.

Second, to investigate whether participants are specifically accurate about their friend, we iteratively permuted participants' dyad assignments. In each iteration, we calculated bivariate correlations between participants' predictions and the actual transitions of their randomly assigned "partner". These correlation coefficients were then Fisher's z-transformed, and their mean was taken. Across 10,000 iterations, we obtained an empirical sampling distribution of mean mismatched person-specific accuracy. If participants' predictions were

specifically accurate for their friend, the correctly-matched mean accuracy from analysis one should occupy a high percentile when compared against the mis-matched accuracy distribution. Finally, we tested whether people make person-specific predictions by invoking person-specific knowledge, or by drawing up either their understanding of general emotion dynamics or knowledge about their own emotion dynamics. To tease apart these sources of knowledge, we used linear mixed effect modeling and model comparison to examine the independent contribution of each of the three ratings (i.e., Friend ratings, Group ratings, and Self ratings) in predicting Friend-Actual. The linear mixed effect models were implemented through the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2020). If participants use person knowledge to tune their predictions of emotion transitions, Friend ratings should predict Friend-Actual even after controlling for Self ratings and Group ratings.

We investigated group-level accuracy in two ways. First, to test how accurately participants can predict an average group member's emotion transitions, we calculated bivariate correlations between the participants' predictions for the group (Group ratings) to predict the group's actual average Self ratings (Group-Actual). These correlation coefficients were then Fisher's z-transformed, and subjected to a one-sample t-test. Second, to tease apart the sources of knowledge that people use to make group-level predictions, we used linear mixed effect modeling and model comparison to examine the independent contribution of Group ratings and Self ratings in predicting Group-Actual. If participants use group knowledge to tune their predictions of emotion transitions, Group ratings should predict Group-Actual even after controlling for Self ratings.

Social success measures. Participants completed measures of two types of social success, general social success and relationship success. General social success was assessed with validated measures of social network size and social support, and loneliness (the UCLA loneliness scale; D. Russell, 1996). Both the size of one's social network (Campbell, Marsden, & Hurlbert, 1986) and one's feeling of loneliness (Gow, Pattie, Whiteman, Whalley, & Deary, 2007) have been used to index how well an individual fares in the general social arena. Relationship success was measured with the Respondent Affection subscale of the McGill Friendship Questionnaire (MFQ; Mendelson & Aboud, 1999) and self-reported closeness. The MFQ is an empirically validated measure of positive feelings in friendships, and has been widely used as a measure of relationship quality (Buote et al., 2007; Mendelson & Kay, 2003). Self-reported closeness is a face valid measure of the subjective sense of interpersonal closeness in a friendship, and has also used in social psychology (Bahns, Crandall, Gillath, & Preacher, 2017) and neuroscience (Welborn & Lieberman, 2014).

To test whether accurate predictions about a friend relate to relationship success, we assessed the relation between participants' person-specific accuracy and their friend's responses on the relationship success measures. To test whether accurate predictions about the group relate to general social functioning, we assessed the relation between participants' group-level accuracy and their own responses on the general social success measures. These analyses were conducted using bivariate Pearson correlations, bootstrapped to obtain 95% confidence intervals for these correlation coefficients.

In addition, participants completed the Reading the Eyes in the Mind test to measure social perceptive ability. The RMET served as a reference point for the validity of person-specific and group-level accuracies as predictors of social success. We assessed the RMET's

relation with measures of both friendship and general social success, as well as both measures of predictive accuracy. Participants completed the social success measures in random order.

Finally, we collected data on participating dyads' friendship duration, and assessed whether it was associated with person-specific accuracy.

1.1.2 RESULTS

PERSON-SPECIFIC ACCURACY. Our primary analyses investigated whether participants could accurately predict the self-reported emotion transitions of a specific target, namely their friend in the current study. To do so, we first calculated the bivariate correlation between Friend ratings and Friend-Actual for each participant. These correlation coefficients were r-to-z transformed and subjected to a one-sample t-test. The mean r-to-z transformed correlation was significantly above zero ($M = 0.68$, average $r = 0.57$, $CI = [0.62, 0.73]$, $d = 2.54$), suggesting that participants were indeed able to accurately predict their friends' emotion transitions (Figure. 1.1).

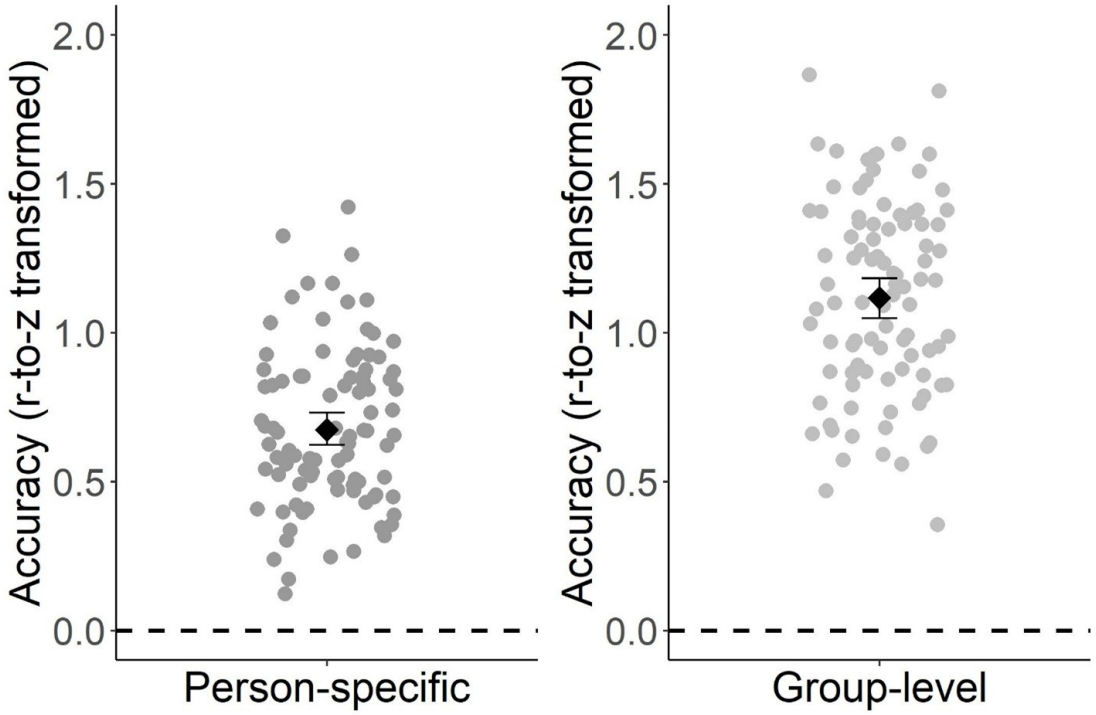

Figure 1.1: Person-specific and Group-level Accuracies. People made highly accurate predictions of the emotion transitions for both their friends (left panel) and average-group member (right panel). Points indicate r-z transformed accuracy for each participant, diamonds indicate means, error bars represent bootstrap 95% confidence intervals, and dashed line represents chance accuracy.

Second, we tested whether participants' predictions are *specifically* accurate for their individual target. To do so, we conducted a permutation test where we calculated how accurate participants' predictions would have been for randomly selected individuals other than their target. We then compared actual accuracy scores against the distribution of these mismatched accuracies. Correctly matched accuracy was significantly greater than chance ($p = .0056$; Figure. 1.2), indicating that participants' person-specific predictions are not only accurate, but also specific to their target (see Appendix A for a conceptually similar analysis).

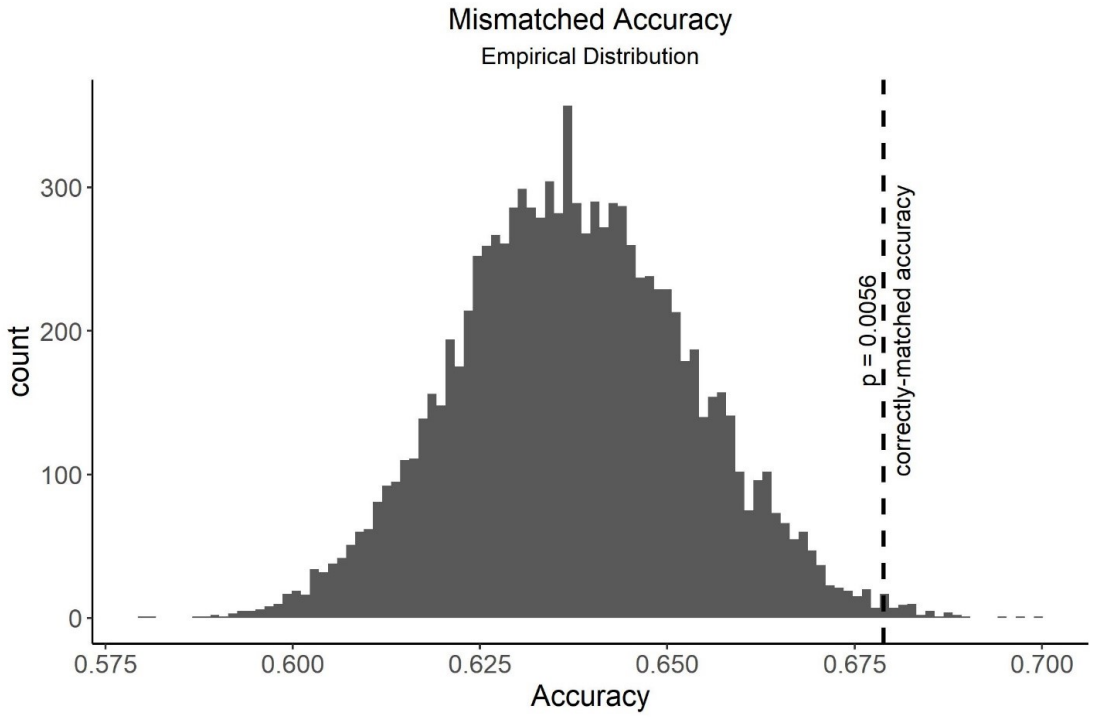

Figure 1.2: Empirical distribution of mismatched accuracy. The distribution was obtained from 10,000 permutations. Dashed line represents the correctly matched person-specific accuracy.

Third, we examined the extent to which people drew upon person-specific knowledge to accurately predicting a friend's emotion transitions. Specifically, we sought to distinguish the contributions of three sources of information to participants' predictions: person-specific knowledge, self-knowledge, and group-level understanding. These contributions were operationalized as the contribution of Friend ratings, Self ratings, and Group ratings, respectively, in predicting Friend-Actual.

Friend-Actual were included as the dependent variable; Friend ratings, Self ratings, and Group ratings were included as predictors. Random intercepts were included for both subjects, nested in dyads, and for items (emotion pairs). The three predictors included in this

model were highly correlated with each other. Self ratings and Friend ratings had a correlation of $r = 0.70$; Group ratings and Friend ratings had a correlation of $r = 0.71$; and Self ratings and Group ratings had a correlation of $r = 0.73$. Thus, the extent to which Friend ratings predicts Friend-Actual in this model reflects how well individuals are able to go beyond their group-level understanding and self-knowledge, and use person-specific knowledge to predict a specific target's emotion transitions.

As expected, results of a linear mixed-effects model showed that participants used tailored, person-specific knowledge about their friend to make their accurate predictions: Friend ratings significantly predicted Friend-Actual ($b = 0.07, \beta = 0.07, t(6887) = 5.60, p < .001$); there was no significant relation between Friend-Actual and neither Group ratings ($b = -0.02, \beta = -0.02, t(6932) = -1.52, p = 0.13$) nor Self ratings ($b = 0.01, \beta = 0.01, t(6920) = 0.49, p = 0.63$). Thus, in spite of a high degree of shared variance, Friend ratings made significant independent contribution to predicting Friend-Actual over and above the contribution of Group ratings and Self ratings. That is, even though participants made largely similar emotion transition predictions for themselves, their group, and their friend, they nevertheless fine-tuned their predictions about a friend using person-specific knowledge.

To further assess the independent contribution of person-specific knowledge to predictive accuracy, we compared the full model above to two reduced models. These model comparisons tested whether Friend ratings was a better predictor of Friend-Actual than Self ratings and Group ratings. The first reduced model used Self ratings and Group ratings (but not Friend ratings) to predict Friend-Actual. The full model performed significantly better than this reduced model ($BIC_{full} = 62890, BIC_{reduced} = 62913, \chi^2 = 31.22, p <$

0.001), indicating that the inclusion of participant's person-specific ratings allowed the model to fit the data significantly better. We also compared the full model to a second reduced model using only Friend ratings to predict Friend-Actual. Results showed that the full model, including Self ratings and Group ratings, did not perform significantly better than the model with Friend ratings alone ($BIC_{reduced} = 62875, \chi^2 = 2.27, p = 0.32$). These results corroborated the findings from the analyses above, and further confirmed our hypothesis that individual possess the ability to utilize individualized information to accurately predict the emotion transitions of specific individuals.

These multiple regression analyses shared similarities with componential approaches to modeling interpersonal accuracy (Furr, 2008). In a componential approach, interpersonal accuracy is modeled as a combination of different accuracy components. One example is the Social Accuracy Model (Biesanz, 2010), where accuracy is jointly determined by components such as normative accuracy, how much one's impressions of others correspond to the characteristics of an average person, and distinctive accuracy, how well one perceives a target's unique characteristics. Analyses conducted in line with the SAM approach replicate results from the multiple regression analysis, showing a significant relation between Friend ratings and Friend-Actual after controlling for Self ratings and Group ratings (see A for details).

Additionally, we found that person-specific accuracy was correlated with friendship duration ($r(92) = 0.16, CI = [0.003, 0.32]$), suggesting that either accuracy promotes durable friendships, or that people learn to become more and more accurate over the course of a friendship.

GROUP-LEVEL ACCURACY. Next, we investigated whether individuals could accurately predict the emotion transitions of the average group member (Group-Actual) using analyses parallel to the person-specific analyses above. First, we calculated the bivariate correlations between Group ratings and Group-Actual for each participant. These correlation coefficients were r-to-z transformed and subjected to a one sample t-test. The mean r-to-z transformed correlation was significantly above zero ($M = 1.12$, average $r = 0.78$, $CI = [1.05, 1.18], d = 3.41$; Figure. 1.1), suggesting that participants accurately retain the aggregate pattern of emotion transitions in their community. This replicated and extended the findings that people can accurately predict population-level transitions (Thornton & Tamir, 2017), here within a personally relevant social group.

Second, we examined the independent contribution of participants' group-level understanding in predicting the aggregate-level emotion transition patterns. We conducted this analysis using a linear mixed effect model. In this model, we investigated the relation between Group ratings and Group-Actual while controlling for Self ratings. Group-Actual and Self ratings were included as predictors whereas Group ratings was included as the outcome variable. Random intercepts were included for participants, nested within dyads, and for items. There was a significant relation between Group ratings and Group-Actual after controlling for Self ratings ($b = 0.63, \beta = 0.49, t(143) = 33.71, p < 0.001$). These results suggest that participants did not simply use their self-knowledge to make predictions about the average group member. Rather, participants have knowledge about the group-level regularities in emotion transitions that contributed independently group-level accuracy.

SOCIAL SUCCESS. To the extent that predicting others' emotions facilitates social interactions, people who can predict emotion transitions with greater accuracy should en-

joy greater social success. First, we tested the extent to which person-specific accuracy correlated with two measures of relationship success (Table 1.2). Participants' person-specific accuracy correlated with their friends' score on the Respondent Affection scale ($r(92) = 0.15, CI = [0.01, 0.30]$). The correlation between participants' person-specific accuracy and their friend's perceived closeness was likewise positive, but not statistically significant ($r(92) = 0.12, CI = [-0.04, 0.25]$). Note that the current sample showed a ceiling effect on the Respondent Affect Scale, with a mean of 2.96 and a median of 3 on a scale ranging from -4 to 4 (see Table 1.1 for means and standard deviations of all social success measures). As such, correlational analyses involving the Respondent Affection scale should be considered with this compression of range in mind. Together, these results suggest that the more accurately a person can predict their friends' emotions, the more successful their friendship.

Next, we tested the extent to which group-level accuracy correlated with general social success (Table 1.2). As expected, participants' group-level accuracy was negatively correlated with loneliness ($r(91) = -0.28, CI = [-0.45, -0.10]$). That is, the more accurately a person can predict their community's emotions, the less lonely they feel. Group-level accuracy was negatively, though not significantly, correlated with participants' social network size ($r(91) = -0.15, CI = [-0.35, 0.05]$) and uncorrelated with participants' support network size ($r(91) = 0.05, CI = [-0.15, 0.25]$). Together, these results provide preliminary evidence that the more accurately a person can predict their group's emotion transitions, the more fulfillment they find in their immediate social environment, though this was not reflected in the quantity of social connections.

Finally, we tested the relation between a traditional measure of social cognitive ability,

the Reading the Mind in the Eyes Test, and each measure of social success (Table 1.2). This serves a reference point for the convergent and discriminant validities of person-specific and group-level accuracies. Participants' RMET was positively associated with their friends' perceived closeness ($r(86) = 0.21, CI = [0.03, 0.38]$); RMET was not correlated with any other measure of friendship or general social success (see Table 1.2). RMET was also uncorrelated with person-specific accuracy (r(86) = 0.08, CI = [-0.12, 0.28]) and group-level accuracy ($r(86) = 0.09, CI = [-0.08, 0.28]$). In contrast, person-specific and group-level accuracy were significantly correlated ($r(85) = 0.50, CI = [0.34, 0.65]$) suggesting common learning or inferential processes might underlie both. Together, these results suggest that the accurate prediction of emotion transitions may be a construct independent from social perceptive ability as measured by the RMET.

	Person-specific Accuracy	Group-level Accuracy	RMET
Relationship success			
Friend's Respondent Affection	0.15*		0.17
	[0.01, 0.30]		[-0.02, 0.33]
Friend's Perceived Closeness	0.12		0.21*
	[-0.04, 0.25]		[0.03, 0.38]
General Social Success			
Social Network Size		-0.15	0.13
		[-0.35, 0.05]	[-0.08, 0.30]
Social Support		0.05	0.05
		[-0.15, 0.25]	[-0.15, 0.23]
Loneliness		-0.28*	-0.08
		[-0.45, -0.10]	[-0.29, 0.12]
Friendship Duration	0.16*		0.07
	[0.003, 0.32]		[-0.16, 0.27]

Table 1.2: Correlations with social success measures from Study 1. Pearson correlations between social success measures and Reading the Mind in the Eyes (RMET), Person-specific Accuracy, and Group-level Accuracy. Bootstrapped 95% confidence intervals are included in brackets. Significance is indicated by asterisks.

1.2 Study 2. Emotion Prediction in New Roommates

Study 1 demonstrated that people can specifically and accurately predict the emotion transitions of particular individuals and social groups. Moreover, prediction accuracy was associated with both their relationship success and their general social success. In Study 2, we aimed to extend these findings beyond the context of highly familiar individuals and social groups, and investigate emotion predictions in a sample of newly acquainted dyads: first-year college roommates. Study 2 was the first time point of an ongoing longitudinal study. Longitudinal analyses were preregistered but do not make up part of the current investigation.

Unlike established close friends, first-year college roommates do not yet have significant stores of knowledge about each other. Studying this population allows us to measure emotion prediction amongst individuals with smaller stores of specialized knowledge. Thus, we can assess how quickly people learn the specialized knowledge necessary to make accurate predictions. By the same token, a first-year college student sample would allow us to investigate the accuracy of group-level predictions in members that are newly introduced to the group, as well as how quickly people can learn to make more accurate group-specific predictions.

In the current study, we make several predictions. First, we predict that the sample of new students tested in Study 2 should be able to capitalize on general regularities in emotion transitions, and achieve accurate predictions both for their new acquaintance and for their social group. However, we also predict that they should be less accurate than the sample of close friends tested in Study 1, as these students had had more time to develop spe-

cialized knowledge about their partner and their social group. We will capitalize on variance in the date of participation across our sample to test for the development of knowledge over time. We expect that the more time that a member of this current sample has spent with their new acquaintance and social group, the more accurately they should be able to predict that acquaintance and social group.

Finally, we test for the relation between predictive accuracy and social success. We expect that this relation emerges slowly over time, as a result of repeated interactions. Recently acquainted dyads and community members should not yet have reaped the benefits that accurate predictions confer on social success. Thus, we do not expect to see a correlation between new students' person-specific accuracy and relationship success, nor between group-level accuracy and general social success.

1.2.1 Methods

PARTICIPANTS. Participants were first-year undergraduate students at Princeton University. Each student signed up together with one roommate. Fifty-nine newly formed roommate dyads completed the study ($N = 118$, 88 women; $M_{age} = 18.44, SD = 1.65$; 46.6% White, 33.9% Asian, 7.6% Black or African American, 8.5% more than one race, 1.7% others, 0.8 % Alaskan Native and American Indian, and 0.8% unreported; 9.3% Hispanic or Latino). Participants were recruited at the very beginning of the school year, within six weeks of the first day of the semester. All students successfully completed in that time frame (mean = 24 days; range = 0-40 days). All roommates were assigned by University housing rather than self-selected. Participants knew each other for 26.7 days on average (range = 0-83 days). One dyad was excluded from analyses due to prior acquaintanceship

(relationship duration = 345 days). No other dyads in the sample were prior acquaintances. Data for Study 2 were collected as the first wave of a longitudinal project. A sample size of 40 dyads was set for the longitudinal study. Fifty-nine dyads were recruited for the first wave to allow for potential attrition.

Emotion transition task. The emotion transition task and the accuracy measures obtained from the task were identical to those used in Study 1.

Social Success Measures Participants completed measures of both Relationship success and General social success. Relationship success was measured using the Reis-Shaver Intimacy Index (Reis & Shaver, 1988), the Respondent Affection subscale of the McGill Friendship Questionnaire (Mendelson & Aboud, 1999), and subjective ratings of closeness and liking used in Study 1. We tested the relation between participants' person-specific accuracy and their roommate's responses on the relationship success measures. General social success was measured using the Social Network Index, the UCLA loneliness scale, and network nomination measures used in Study 1. We tested the relation between participants group-level accuracy and their own responses on the general social success measures. All relations were tested using the same procedure as Study 1.

Finally, participants completed the Reading the Eyes in the Mind test. The RMET was used to test the convergent and discriminant validities of both person-specific and group-level prediction accuracy measures. We assessed the RMET's relation with relationship and general social success, as well as both measures of predictive accuracy. Participants completed all measures in random order.

1.2.2 Results

Person-specific Accuracy .We investigated whether participants could accurately predict their roommate's self-reported emotion transitions in three analyses. First, we calculated the bivariate correlation between Friend ratings and Friend-Actual for each participant. These correlation coefficients were then subjected to a one-sample t-test after r-to-z transformation. The mean r-to-z transformed correlation was significantly above zero ($M = 0.51$, average $r = 0.44$, $CI = [0.46, 0.57]$, $d = 1.69$), evidencing that members of newly formed dyads can already accurately predict their target's specific patterns of emotion transitions (Figure 1.3). However, this accuracy is significantly lower than that observed in established close friends dyads in Study 1 ($F(1, 208) = 17.99, p < .0001$).

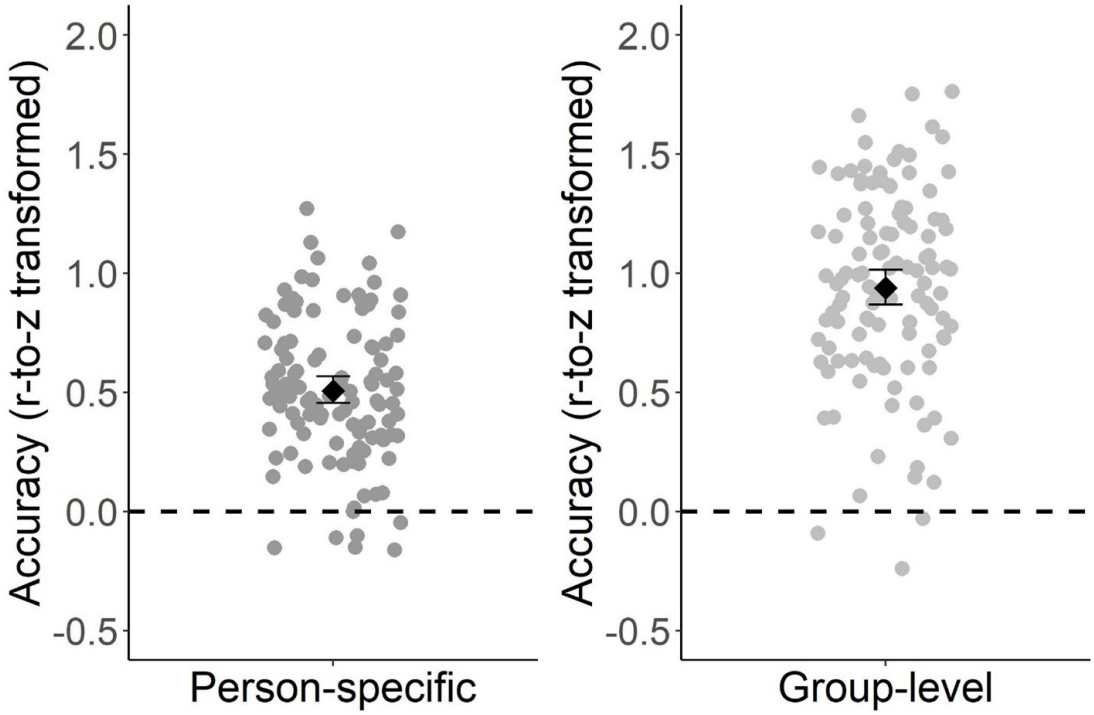

Figure 1.3: Person-specific and Group-level Accuracies form Study 2. People made significantly accurate predictions of the emotion transitions of both their roommate (left panel) and average-group member (right panel). Points indicate r-z transformed accuracy for each participant, diamonds indicate means, error bars represent bootstrap 95% confidence intervals, and dashed line represents chance accuracy.

Second, we tested whether participants' predictions were specifically accurate to their individual target. We iteratively permuted participants' dyad assignment to obtain an empirical sampling distribution of mismatched accuracy. Correctly matched accuracy was significantly greater than chance ($p = .0016$; Figure 1.4), indicating new college roommates can already make person-specific predictions.

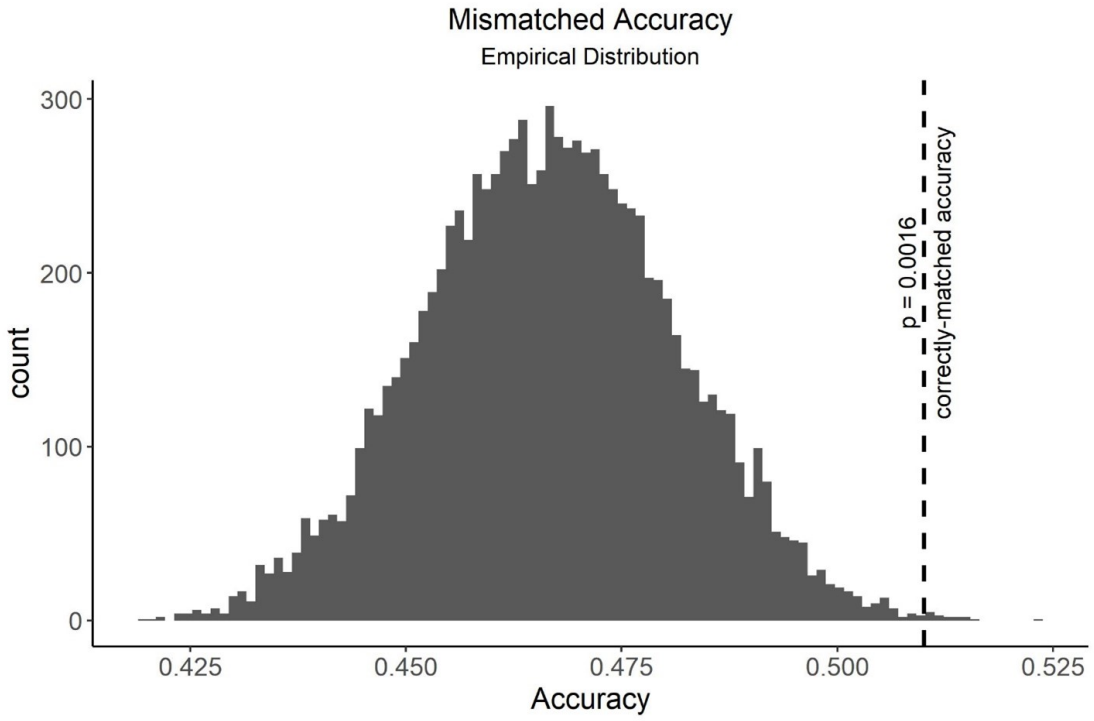

Figure 1.4: Empirical distribution of mismatched accuracy. The distribution was obtained from 10000 permutations. Dashed line represents the correctly matched person-specific accuracy.

Third, we assessed the contribution of person-specific knowledge in accurately predicting specific targets' emotion transitions, using the same analysis as in Study 1, with a linear mixed effect model predicting target's Self ratings with a participants' Friend ratings, Self ratings, and Group ratings. Results indicated that tailored person knowledge contributed to accurate person-specific predictions. Friend ratings significantly predicted Friend-Actual after controlling for Self ratings and Group ratings ($b = 0.05, \beta = 0.05, t(8654) = 4.54, p < .001$). Neither Group ratings ($b = 0.01, \beta = 0.01, t(8633) = 0.85, p = .40$) nor Self ratings ($b = 0.02, \beta = 0.02, t(8655) = 1.31, p = .19$) significantly predicted Friend-Actual in this model. Consistent with results from Study 1, participants made sim-

ilar predictions of emotion transitions for themselves, their group, and their friend. Self ratings and Friend ratings had a correlation of $r = 0.62$; Group ratings and Friend ratings had a correlation of $r = 0.64$; and Self ratings and Group ratings had a correlation of $r = 0.68$. Nonetheless, Friend ratings made significant independent contribution to predicting Friend-Actual, indicating that even participants in new relationships fine-tuned their person-specific predictions using individualized knowledge.

To further validate the results of the linear mixed effects model, we tested whether Friend ratings better predicted Friend-Actual than did Self ratings and Group ratings. We did so by comparing the full model above to two reduced linear mixed effect models. The first reduced model used Self ratings and Group ratings (but not Friend ratings to predict Friend-Actual). The full model performed significantly better than this reduced model ($BIC_{full} = 78435, BIC_{reduced} = 78446, \chi^2 = 20.54, p < .001$). This indicates that including person-specific ratings allowed the model to fit the data significantly better. The second reduced model used only Friend ratings to predict Friend-Actual. The full model, even though including Self ratings and Group ratings as additional predictors, did not significantly outperform this second reduced model ($BIC_{full} = 78435, BIC_{reduced} = 78420, \chi^2 = 3.51, p = .17$). Model comparison results corroborated our hypothesis that accurate person-specific predictions are made by incorporating individualized knowledge about the target person.

The results thus far suggest that participants are able make accurate and specific predictions about a new roommate, within only a few weeks of acquaintance. Thus, we next assessed whether people accumulate this person-specific knowledge over time. That is, we tested the bivariate correlation between person-specific accuracy and relationship dura-

tion. Participants' person-specific accuracy for their roommate was significantly correlated with the number of days since they first met when the study was completed ($r(114) = 0.23, CI = [0.07, 0.37]$). Even at a very early stage of relationship, dyads that have known each other longer can more accurately predict each other's emotion transitions. Since relationship duration cannot be a product of predictive accuracy in new roommates, this result potentially suggests that people can rapidly acquire information that is needed to accurately predict specific individuals' emotion transition. To estimate participants' accuracy at zero acquaintance, we conducted an exploratory linear model equivalent to this bivariate test, using relationship duration to predict person-specific accuracy. This model had an estimated intercept of .38. That is, at zero acquaintance, people can already somewhat accurately predict their roommate's emotion dynamic, possibly by relying on their general understanding. Nonetheless, they quickly gather necessary knowledge, and finetune their person-specific predictions to become more accurate. To estimate the speed with which people acquire the information necessary to make accurate predictions at the level of a close friend, we conducted an exploratory extrapolation of this model. This analysis suggested that, under the assumption of linearity, participants need approximately 61 days to reach the accuracy level of close friends.

GROUP-LEVEL ACCURACY. We next tested whether individuals could accurately predict the average group member's emotion transitions, using the same analyses as in Study 1. We first conducted a one sample t-test on r-to-z transformed, per-subject correlations between Group ratings and Group-Actual. The mean r-to-z transformed correlation was significantly above zero ($M = 0.94$, average $r = 0.68, CI = [0.87, 1.02], d = 2.31$; Figure 1.3), indicating that participants have an accurate understanding of the aggregate patterns of

emotion transitions in their new group. However, group-level accuracy from new college students is significantly lower than its counterpart obtained from the older college students in Study 1 ($F(1, 207) = 11.59, p < .001$).

We also examined the contribution of participants' group-level understanding to predicting the aggregate emotion transition patterns. In a linear mixed effects model, we used Group-Actual and Self ratings to predict Group ratings. Random intercepts were included for subjects nested in dyads, as well as for items (emotion pairs). A significant relationship between Group-Actual and Group ratings in this model would indicate that group-level understanding independently contributes to accurate aggregate level predictions above and beyond participants' self-knowledge. Indeed, we found that Group-ratings significantly predicted Group ratings after controlling for Self ratings ($b = 0.05, \beta = 0.05, t(8654) = 4.54, p < .001$).

These results indicate that participants can make accurate predictions about group-level transitions within only a few weeks of joining a new community. We next examined whether people accumulate this group-level knowledge over time. We tested the relationship between duration of group membership and group-level accuracy. How well participants can predict an average group member's emotion transitions was positively correlated with the number of days the student had been on campus ($r(114) = 0.20, CI = [0.03, 0.36]$), suggesting that people can quickly learn group-level emotion dynamics, even over a short period of time.

To estimate participants' accuracy at zero acquaintance with the group, we conducted an exploratory linear model equivalent to this bivariate test, using number of days to predict group-level accuracy. The model had an estimated intercept of .75. That is, without having

spent any time in the group, people can already somewhat accurately predict the group's emotion dynamics, possibly by relying on their general understanding. Nonetheless, they quickly fine-tune their group-level predictions to become more accurate. To estimate the speed with which people acquire the information necessary to achieve the level of accuracy of a longstanding community member, we conducted an exploratory extrapolation of this model. This analysis suggested that, under the assumption of linearity, participants need approximately 46 days to make accurate predictions at the level of longstanding community members.

Social Success. Using bivariate correlations, we first tested the relation between person-specific accuracy and measures of relationship success (1.3). Person-specific accuracy was not significantly correlated with the roommates' responses on the Respondent Affection scale ($r(114) = 0.11, CI = [-0.07, 0.28]$), the Reis-Shaver Intimacy model ($r(114) = 0.17, CI = [-0.004, 0.34]$), or perceived closeness ($r(114) = 0.02, CI = [-0.16, 0.21]$). As expected, in the current sample of new college roommates, more accurate prediction of emotion transitions did not yet translate into or reflect the quality of roommate relationships. Nonetheless, these results are directionally consistent with results from Study 1, suggesting that the positive relation between person-specific accuracy and relationship success might develop over time.

We next assessed the relation between group-level accuracy and general social success in the current new college student sample. The more accurately a student predicted the group's emotion transitions, the larger their social network size ($r(114) = 0.29, CI = [0.14, 0.43]$) and support network size ($r(114) = 0.24, CI = [0.07, 0.39]$). Group-level accuracy was negatively associated with loneliness ($r(114) = -0.13$) as in Study 1, though

not statistically significantly so ($CI = [-0.30, 0.04]$). The relation between group-level accuracy and loneliness was stronger in Study 1 than in Study 2, which is consistent with the hypothesis that the relation between group-level accuracy and general social success develops over time. On the other hand, the relation between group-level accuracy and social network measures was non-significant in Study 1, but significant in Study 2. In all likelihood measures of social and support network reflect relationships prior to entering college. As such, their relationship with group-level accuracy should be interpreted with this caveat in mind.

Finally, we tested the extent to which measures of person-specific and group-level accuracy converge with the RMET. We found that participants' score on the RMET was strongly correlated with both their person-specific ($r(114) = 0.34, CI = [0.11, 0.52]$) and general ($r(114) = 0.43, CI = [0.22, 0.60]$) accuracy. These findings deviated from the results of Study 1, where we found no relation between RMET scores and predictive accuracy. This result suggests that at an early stage of dyadic relationship, accurate prediction of one's partner's emotion transitions might partially depend on accurate perceptions of their emotional states. However, RMET did not translate into social success in this sample (1.3). RMET was not associated with any measures of relationship success between roommates; RMET was positively associated with social network size ($r(114) = 0.21, CI = [0.02, 0.39]$), but not support network size or loneliness.

	Person-specific Accuracy		RMET	RMET
Relationship Success				
Roommate's Respondent Affection	0.11		-0.07	0.17
	[-0.07, 0.28]		[-0.27, 0.10]	[-0.02, 0.33]
Roommate's Perceived Closeness	0.02		-0.13	0.21*
	[-0.16, 0.21]		[-0.31, 0.05]	[0.03, 0.38]
Roommate's Reis Shaver Intimacy	0.17		-0.03	
	[-0.004, 0.34]		[-0.22, 0.14]	0.13
General Social Success				[-0.08, 0.30]
Social Network Size			0.21*	0.05
			[0.02, 0.37]	[-0.15, 0.23]
Social Support		0.24*		-0.08
		[0.07, 0.39]		[-0.29, 0.12]
Loneliness			-0.04	0.07
			[-0.18, 0.12]	[-0.16, 0.27]
Relationship Duration	0.23*		0.11	
	[0.07, 0.37]		[-0.04, 0.25]	
Days as group member		0.20*	0.04	
		[0.03, 0.36]	[-0.13, 0.21]	

Table 1.3: Correlations with social success measures from Study 2. Pearson correlations between social success measures and Reading the Mind in the Eyes (RMET), the Empathic Concerns and the Perspective Taking subscales of the Interpersonal Reactivity Index, Person-specific Accuracy, and Group-level Accuracy. Bootstrapped 95% confidence intervals are included in brackets. Significance is indicated by asterisks.

1.3 Discussion

People's thoughts and feelings unfold over time. These emotion dynamics often follow predictable patterns –happiness is more likely to follow awe than anger; clarity is more likely to follow thinking than pride. However, emotion dynamics are also idiosyncratic. They can vary widely from person to person and from group to group. The current set of studies demonstrates that people can make accurate, fine-tuned predictions of emotion transitions for both individuals and groups. These findings thus highlight a novel set of social cognitive skills, namely social predictions, as well as the implications of these skills for social success.

We found evidence that people can make accurate predictions of emotion transitions in two very different type of dyads: close friends, and new roommates. Previous work has demonstrated that people can make accurate judgments about specific, well-known others in domains such as mental content (Stinson & Ickes, 1992) and behavior frequency (Vazire & Mehl, 2008). Our results extend these general findings about accurate social inferences into the domain of social prediction, namely the prediction of others' future states. Further, we provide evidence for how people accomplish this mind reading feat: people tailor their predictions about others by drawing on knowledge about their specific target. That is, people do not blindly apply their self-knowledge or their general understanding of emotion transitions to make person-specific predictions. While people understand that, in general, emotion dynamics follow predictable patterns, they also recognize the idiosyncrasies in their targets' dynamics. They fine-tune their predictions by taking these idiosyncrasies into account. The idea that people use person-specific knowledge to calibrate social predictions

dovetails nicely with the previous finding that increased levels of empathic accuracy between male friends, as compared to between strangers, was accounted for by the detailed knowledge friends had about each other (Stinson & Ickes, 1992).

In addition, we found that it takes time for people to accumulate target-specific knowledge. This is evident on two different timescales. First, we can look within the sample of new students. Students in this sample rapidly accrued the knowledge they needed in order to make accurate predictions about both their roommate and their community. Students who knew their roommate for only a few weeks were already much better at predicting their roommate than people who knew their roommate for only a few days. However, a few weeks is not nearly enough time to learn all one needs to know in order to make accurate predictions. We can see the effect of additional time by comparing our sample of new roommates to our sample of established friends. New dyads were significantly less accurate than dyads in longstanding relationships. We found similar effects when looking at changes in accuracy about the group as well. People in new communities can already make accurate predictions about an average group member, but they do better and better with more exposure to the group, and longstanding group members outperform newer ones. People gradually acquire the knowledge that they need, and become more accurate in their predictions over time. These findings highlight the importance of target-specific knowledge for fine-tuning social predictions.

We found evidence that people can accurately predict emotion transitions for their local community. We suggest that people solve this group prediction problem by accumulating information about their specific social group to gain a more accurate understanding of group-level emotion dynamics. Indeed, first-year students were less accurate than up-

perclassman students who had more time to accumulate group knowledge; and first-year students who participated shortly after arriving on campus were less accurate than students who had been on campus for longer. That said, we did not measure how accurate people were about any broader social group (e.g., Americans, people in general), so it is possible that people were able to be accurate about their local group by simply applying general knowledge in making group-level predictions. Future research should aim to provide direct evidence of group specificity by collecting predictions at both the community level and the broader group level.

Our results raise important questions about the process by which people learn to make person- and group-specific predictions. People may learn to make target-specific predictions in at least two ways. One possibility is that people learn incrementally through direct observations. People might observe and accumulate many individual instances of emotion transition over time, and then construct their target-specific mental models of emotion dynamics by simple tallying. Alternatively, target-specific learning might be scaffolded by prior beliefs over how different emotion transitions tend to co-occur. Guided by such abstract knowledge, people could make inductive inferences about a specific target's unobserved emotion transitions after accumulating a small number of observations. The two hypothesized learning processes make different predictions about how long the learning process might take; the former requires a large amount of input that could only be acquired over a long period of time, whereas the latter requires only a few observations before generalization is possible, and would be much less time consuming. Here, we find that even within the very early stage of relationship formation, those who had interacted longer before completing the study tended to be more accurate. This finding favors the hypothesis

that people rely on inductive inferences to quickly hone their target-specific mental models, rather than building these models from scratch. Future research should further investigate how abstract knowledge and inductive inferences facilitate learning about specific individual targets.

Accurate social predictions rely on target-specific knowledge. As a result, predictive accuracy should be partially dissociable from more traditional constructs of social cognition, such as social perception. We found that accurate person-specific predictions were highly correlated with the RMET in new dyads, whereas the same association was much weaker in established dyads. Local perceptual information, such as facial expressions and tones of voice, might serve as the most available and effective input about others' internal states at early stages of relationship formation. That is, the ability to infer mental experiences from the perceptible might be important for social predictions early on. As relationships develops, however, people need to go beyond the immediately observable in order to refine their understanding of a target's emotion dynamics. This further refinement might require the additional ability to learn and represent statistical contingencies over a longer timescale (Buchsbaum, Griffiths, Plunkett, Gopnik, & Baldwin, 2015), or the ability to make complex social and causal attributions (Alicke, Mandel, Hilton, Gerstenberg, & Lagnado, 2015). How well people fare in their social lives likely depends not only on how well they can reactively process indications about others' mental states that are already in place, but also proactively form expectations about how others might think, feel, and act in the future. Future research should attempt to further investigate the extent to which social predictions might be a novel domain of social cognition using more comprehensive approaches.

Our findings offer preliminary evidence that social predictions bestow important social advantages. Long-term close friends that more accurately predict each other's emotion dynamics enjoy better relationship, and long-term community members that more accurately grasp their group's emotion dynamics enjoy better general social wellbeing. These findings converge with previous literature on how social cognitive abilities positively relate to social well-being and social success (Banerjee et al., 2011; Gleason et al., 2009; Sened et al., 2017; Morelli, Ong, Makati, Jackson, & Zaki, 2017; Bosacki & Astington, 1999). However, this relation between accurate predictions and social success did not hold amongst social agents newly introduced to a relationship and a community. The relationship between accurate predictions and social success likely emerges over time. Interestingly, this emerging process seems to take place quickly in the early courses of relationship formation. Exploratory analyses suggest that the association between person-specific accuracy and relationship success was stronger among students who completed the study later on during recruitment (see A). This result provides evidence that the interplay between accuracy and social success might start occurring shortly after the social context forms, and the dividends of learning to make accurate predictions might accrue over time. However, there are several caveats to note about this finding. First, the correlations between predictive accuracy and social success, while predicted a priori, have small effect sizes in Study 1. Second, while we hypothesized a priori that these relations should not exist in Study 2, it is also possible that we were simply unable to detect a relation between predictive accuracy and social success due to our sample size. Given the small effect size of in Study 1, we might not be well-powered enough to detect these relations in Study 2, especially if they were already attenuated. Third, the relation between accuracy and social success did not hold for all measures of social success.

Fourth, the cross-sectional nature of our comparison also does not allow us to eliminate the influence of self-selection, and thus limits the strength of our claim that the benefit of social predictions accrues over time. Thus, our results provide only preliminary evidence for a relation between predictive accuracy and social success. Future investigations should aim to replicate these results at a larger scale, in order to achieve the statistical powers needed to establish both the positive and the null findings with higher confidence.

If it is the case that social prediction enables social success, then we can further ask two questions about the nature of this relationship. First, does social prediction contribute more to social success than other components of social cognition, such as social perception? In the current sample, predictive accuracy was associated with social success at least to the same extent as, if not better than, did the RMET. Social relationships are interpersonal in nature (Schilbach et al., 2013). Social accuracy in most situations can be difficult to objectively assess, and within-person, in-lab measures, such as emotion recognition, need not align with how well a perceiver predicts an actual target's emotions (Zaki & Ochsner, 2011). In contrast, predictions of emotion transitions can be measured against empirically obtained benchmarks (Thornton & Tamir, 2017). Social prediction, as operationalized in the current study, is similarly a purposeful measure of accuracy. It is thus a useful assay of real-world social ability precisely because it aims to map real-world experiences. It is worth noting that the current study measured accuracy against self-reported experiences, rather than experience-sampled ones. Nevertheless, this study found similar levels of accuracy ($\rho_{specific} = 0.56, \rho_{group} = 0.76$) as three studies with experience-sampled benchmarks ($\rho s = 0.77, 0.68, and 0.79$; Thornton & Tamir, 2017). Nonetheless, future research should consider using empirical benchmarks to measure person-specific accuracy, to fully cap-

ture the relation between social prediction and social success. Future work should also try to further establish the convergent and discriminant validities of social prediction against more ecologically sound, open-ended measures of conventional social cognitive ability (Betz, Hoemann, & Barrett, 2019; Cassels & Birch, 2014).

Second, to what extent does predictive accuracy actually cause social success? Our results are consistent with the hypothesis that accurate social prediction leads to social success. However, it is equally possible that people in better relationships are simply more motivated to be accurate in their predictions about their target, or that a third variable causes both. Similarly, while it is possible that higher group-level accuracy leads to higher general social success, it is also possible that frequent exposure to the social group underlies both accuracy and social success. The cross-section and correlational nature of the current study allows us to draw only correlational conclusions about the relation between social predictions and social success. Future studies could resolve this issue by employing a longitudinal design, and tracking how predictive accuracy and social success both develop over time. A longitudinal design would allow researchers to make stronger directional claims about the relation between predictive accuracy and social success. Furthermore, a longitudinal design would allow, researchers to investigate how other variables known to impact relationship formation, such as perceived and objective similarity between the dyad members (Bahns et al., 2017), might impact and interact with predictive accuracy in a developing relationship.

Finally, it is important to note that interpersonal accuracy can be conceptualized and measured in various ways. Accuracy in the current work, in the domain of predicting emotions, has been indexed using correlational metrics between a person's prediction and either self-reported or experience-sampled ground truths. A similar correlational approach is also

taken in the SAM (Biesanz, 2010). This approach conceptualizes accuracy as the relative correspondence between two measures across a wide range of items. However, accuracy can also be conceptualized as the absolute numerical agreement between measures. One prominent example of this approach can be found in the affective forecasting literature (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; Wilson & Gilbert, 2003), which finds biases in people's predictions by looking at deviations in the magnitude of a predicted emotion on a smaller number of emotions. Future studies should combine more complex designs, such as those containing network structures, with varying analytical approaches like the SAM, in order to gain a more nuanced understanding of how people accurately predict others' emotions.

Navigating daily social life is one of human beings' crowning achievements. The social world demands that we constantly interact with a wide array of individuals, each with their own complex characteristics, and each tied to us in different ways. Most people expertly handle these challenges. The present study highlights that people can make accurate predictions about others' future emotions at multiple levels of specificity, ranging from well-known friends to personally meaningful social groups. Our findings add a new layer to how powerfully and flexibly our cognitive system deals with demanding social tasks, but also indicate that there is still much to explore.

2

Rational Egocentric Projection for Accurate Emotion Prediction

To have positive, successful social interactions, people must accomplish the key task of social prediction. They must anticipate how others might think, feel, and act in the future, and ready themselves accordingly. The ability to see into the social future is particularly important when people are trying to establish new interpersonal connec-

tions. The stakes are high with new social targets, as first impressions are influential and long lasting (Rosenbaum, 1986). Moreover, the task itself is more challenging. When making predictions about new acquaintances, people have no shared history or accumulated person-specific knowledge to turn to. Instead, they must rely on other social prediction strategies. How do people resolve the challenge of social prediction, even in the absence of rich knowledge about the target?

Egocentrism –anchoring social inferences on self-knowledge –is one default strategy in the social prediction toolkit. It is a heuristic solution that is often seen as error prone. However, we propose that egocentrism is a rational decision strategy, for at least two reasons. First, we hypothesize that egocentrism generates accurate predictions. Second, we hypothesize that people deploy it strategically, based on how representative they are of their social environment. In two studies, we set out to test these interrelated ideas.

2.0.1 Social Prediction Facilitates Social Connections

People have a fundamental need to create meaningful social relationships (Baumeister & Leary, 2017). Throughout their lives, people strive to connect with the people they encounter for collaboration, friendship, and romantic intimacy. To form these connections, it is crucial that people foster positive interaction with others, especially in the early stages of acquaintanceship.

One critical skill that enables successful social interactions is the ability to see into the social future (Tamir & Thornton, 2018). That is, social perceivers need to anticipate how others might think, feel, and even act before these events take place. By doing so, people can prepare their own actions ahead of time, and thus make social interactions more seamless.

Oftentimes, people are familiar with the targets of their social prediction. Through their experiences with familiar targets, people accrue rich, person-specific knowledge that aid them in making predictions (Stinson & Ickes, 1992; Biesanz, West, & Millevoi, 2007; Zhao et al., 2020). Information about the target's trait-like characteristics, as well as how they felt or acted in similar situations in the past, can help greatly in honing social predictions to be more accurate. However, when people make social prediction about novel, unfamiliar targets, as is often the case when they attempt to build new connections, they do not have detailed person knowledge at their disposal. How do people resolve the complex challenge of social prediction when they're confronted with its full force?

One default heuristic used in social prediction is egocentrism (Epley, Keysar, Van Boven, & Gilovich, 2004). In egocentric inferences, people use their rich and easily accessible first-person knowledge –how they themselves might think or feel if they were in their target's situation –to approximate the possible experiences of the target. That is, people ignore a large number of potential cues about the target, thus relaxing the computational challenge of social prediction. Evidence of egocentrism abounds across many domains of social judgment. When asked what others notice or pay attention to, people routinely base their judgments on what is salient to themselves, ascribing their own attentional states to the target (Gilovich, Medvec, & Savitsky, 2000). Along a similar vein, when people are asked about others' knowledge state, they often anchor on what they themselves know. People even assume that private information, such as one's own feelings or intentions, is accessible to other people (Gilovich, Savitsky, & Medvec, 1998; Keysar, 1994; Keysar & Henly, 2002). People rely heavily on self-knowledge extensively in their social judgments.

2.0.2 The Rationality of Egocentrism

Much past research has regarded this egocentric strategy as an irrational misstep. For example, in the seminal work on the *false consensus effect*, participants routinely overestimated how much others shared their opinions and preferences, even when in the minority (Ross et al., 1977). Similarly, research has demonstrated that people tend to ascribe their own mental states to others more than is warranted, leading to the *illusion of transparency* or the *curse of knowledge* (Gilovich et al., 1998; Camerer, Loewenstein, & Weber, 1989; Birch & Bloom, 2007). Research has also focused on the detrimental downstream consequences of egocentrism. For example, an undue sense of self-other agreement has been linked to negative outcomes such as mis-identifying attitudinal norms and undesirable behaviors (Bauman & Geher, 2002), as well as having more polarized attitudes around social issues (Judd & Johnson, 1981; Wojcieszak, 2008). Such a dominant bias framing masks the potential utility of egocentrism.

We suggest that egocentrism can, in fact, be considered a rational strategy for social prediction, for at least three reasons (Lieder & Griffiths, 2020). First, egocentrism is statistically sound in certain circumstances. Specifically, egocentrism reflects a valid statistical procedure for making social predictions if knowledge about the self is the only available piece of evidence (Dawes, 1989). While self-knowledge constitutes a very small sample of all possible data points, it is procedurally rational to integrate this sample into one's prediction in the absence of other information (e.g., information about the target).

Second, egocentrism is effective. That is, self-knowledge can lead to accurate social inferences. Social judgment can be thought of as an optimization problem. To solve this problem, people use knowledge from different sources, such as knowledge about the self

and about different social groups. The challenge is to find the most optimal way to weight and combine these sources of knowledge to make the most accurate judgment possible. How much weight on self-knowledge is optimal? One study calculated participants' optimal level of egocentrism (Hoch, 1987). In this study, participants predicted how much a social target would agree with a series of attitude statements (e.g., "communism is the world's greatest peril"), and reported their own attitudes for the same items. Analyses under this optimization framework showed that the majority of participants would have been more accurate had they been even more egocentric than they actually were. That is, self-knowledge is an effective proxy for knowledge about others, further supporting the value of egocentrism in social inferences.

Third, other sources of information that people rely on during social inference, such as stereotypes, can be inaccurate (Eyal, Steffel, & Epley, 2018; McConnell, Sherman, & Hamilton, 1994; Terracciano et al., 2005). In one series of studies, when participants were encouraged to abandon their egocentric defaults and actively engage in perspective taking by "putting themselves in others' shoes", the accuracy of their interpersonal judgments was found to decrease. These negative effects of perspective taking were attributed to the inaccuracy of participants' stereotypes about groups. This further highlights the potential importance of self-knowledge for accurate social judgment.

In the current study, we test the hypothesis that egocentrism is a rational strategy for social judgments in the domain of emotion prediction. We expect that the more people rely on self-knowledge to make social inferences, the more accurate they will be in their social predictions. Importantly, we test this in the context of real-world social inferences at the very start of new relationships.

2.0.3 Similarity as a Rational Cue to Be Egocentric

To make rational decisions, people must flexibly adapt their strategy based on relevant contextual cues (Griffiths, 2020; Griffiths, Lieder, & Goodman, 2015). When contextual cues indicate that one strategy is more likely to lead to accurate judgment, people should be more likely to employ that strategy. One central cue for egocentrism in the domain of social prediction is dyadic similarity, or the similarity between a perceiver and the target of prediction. When people base their predictions on self-knowledge, their accuracy should increase, but only to the extent that this self-knowledge is also true of their target. Thus, perceivers who are similar to their target should experience the greatest benefit from egocentrism.

Indeed, research has shown that people do not employ an egocentric strategy in a one-size-fit-all fashion. Rather, people selectively engage in egocentrism only when it is most rational to do so: people are more likely to use self-knowledge to make an inference about a similar target than a dissimilar target (Robbins & Krueger, 2005). The effect of similarity on this social inference strategy is both pervasive and profound. Egocentrism reliably increases with interpersonal similarity, including similarity in personality traits (Cho & Knowles, 2013), academic and professional affiliations (Holtz, 2003; Ames, 2004a, 2004b), nationality (Li & Hong, 2016), political orientation (Tamir & Mitchell, 2013), and ethnicity (Granberg, Jefferson, Brent, & King, 1981).

Similarity along one social dimension can affect how much people recruit self-knowledge to make judgments on other, potentially unrelated dimensions. For example, in Tamir and Mitchell (2013), participants engaged in more egocentric strategies when they judged a similar target's habits and preferences, even though target similarity was manipulated using

political beliefs. In studies using a minimal group paradigm (Tajfel, 1970), researchers manipulate perceived similarity by dividing participants into categories using fictitious and uninformative criteria. This design removes any potential correspondence between the similarity dimension and the judgment dimension. Even so, people still employ self-knowledge in their inferences about minimal groups (Clement & Krueger, 2002). People assume that those who are similar to them in some respect will also be similar to them in others, allowing perceived similarity to exert profound influence on social judgments.

If egocentrism leads to overall more accurate social judgments and people flexibly employ self-knowledge when the target is similar, people should make more accurate inferences about similar targets. Indeed, this prediction has been supported empirically. In one study, participants predicted the attitudes of different targets (Hoch, 1987). Those who made predictions about a more similar target (e.g., their spouse or an average graduate school peer) were more egocentric and more accurate than those who made predictions about a dissimilar target (e.g., the average American consumer). In a more recent study (Woo & Mitchell, 2020), participants predicted social targets' attitudes and preferences. They similarly found that when people made predictions about targets that they considered more similar to themselves, their predictions were more egocentric and more accurate. In sum, people rationally adjust their egocentric strategy, ultimately leading to more accurate predictions about more similar others.

2.0.4 Typicality as a Rational Cue to Be Egocentric

Dyadic similarity provides a strong cue to use egocentric prediction. However, it might not be the most useful cue when people make predictions about a new, unfamiliar target.

Dyadic similarity is variable and must be calculated on a case by case basis as people meet new targets. When engaging with a new person, people might not have enough time to comprehensively assess how alike they are. If people adjust their egocentric strategy to an unreliable sense of dyadic similarity, they might make less accurate social predictions.

Instead, we propose that interpersonal similarity informs social inference strategies over a broader timescale. People differ in how similar they tend to be, in general, to others in their immediate social environment. In other words, people are typical to varying extents. Much like similarity at the dyad level, a person's typicality influences the effectiveness of egocentrism. If a person is more typical, they are, by definition, more similar to more people in their social environment. If a person is similar to an individual target, they can make more accurate predictions by relying on self-knowledge in that specific instance. As such, typical people should, on average, make more accurate predictions by being egocentric. Over the course of people's social lives, typical people should gradually learn that they can make more accurate predictions about a novel target simply by defaulting to using themselves as an anchor. If people rationally adapt their inferential strategy in this way, we should see more typical people use more self-knowledge to make social prediction.

Prior work on typicality has primarily focused on the typicality of a perceived object, rather than the perceiver (Cantor & Mischel, 1979; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Nosofsky, 1988; Rosch, Simpson, & Miller, 1976). Extant studies on the typicality of a perceiver have largely investigated the agent's own perceived typicality. For example, this work finds that a group members' own perceived typicality can predict how likely they will be to "go to the extreme" for the group (Goldman & Hogg, 2016), or to align themselves with stigmatized groups (Essien, Otten, & Degner, 2020). In the cur-

rent study, we focus not on how typical one perceives themself to be. Rather, we focus on ground truth typicality –how typical people actually are in relation to their social surroundings. Self-knowledge can only be a valid cue for accurate social prediction to the extent that people are actually typical.

2.0.5 The Current Studies

Across two studies, we aimed to demonstrate and replicate, respectively, that (i) egocentrism is a rational strategy that leads to accurate predictions, and (ii) perceivers rationally adapt their egocentrism based on their own typicality and dyadic similarity. To do so, we investigate three sets of interrelated hypotheses: First, we hypothesize that egocentrism is an overall effective strategy for social prediction. We expect that greater egocentrism should be associated with more accurate social predictions.

Second, we hypothesize that perceiver typicality is an informative cue for using egocentric strategies. We first replicate prior work showing that dyadic similarity leads people to rely on self-knowledge. We propose that individuals develop their social inference strategies over a lifetime, such that the typicality of a perceiver serves as the most informative cue for employing egocentric strategies. We expect that more typical perceivers should use more egocentric strategies.

Finally, we hypothesize that perceiver typicality improves how accurately people make predictions about novel individuals. Specifically, it does so by increasing the use of egocentric social inference strategies between the perceiver and their target. That is, we expect more typical perceivers will be more egocentric, which in turn leads to more accurate predictions.

2.1 Study 3. Rational Egocentrism for Accurate Emotion Prediction

2.1.1 Methods

participants. Participants were first-year undergraduate students at Princeton University. Each student signed up for the study with one roommate at the outset of the fall semester. Participants were recruited as the first wave of a longitudinal study, though subsequent waves are not analyzed as part of this investigation. A sample size of 40 dyads was set a priori, and 59 newly formed roommate dyads were recruited to allow for attrition (N = 118, 88 women; Mean age = 18.44, SD = 1.65; 46.6% White, 33.9% Asian, 8.5% more than one ethnicity, 7.6% Black or African American, 1.7% others, 0.8% Alaskan Native and American Indian, and 0.8% unreported; 9.3% Hispanic or Latinx).

All participants were recruited within six weeks of the first day of the semester, and all completed the study within that time frame (mean = 24 days; range = 0-40 days). Participants were assigned roommates by the university and had known each other for an average of 26.7 days (range = 0-83 days). One dyad was excluded from analyses because of prior acquaintanceship (relationship duration = 345 days). All participants provided informed consent in a manner approved by the Princeton University Institutional Review Board.

procedures. Participants were recruited in pairs but instructed to complete the study online independently. During the study, participants first completed an emotion transition task in which they made predictions about their own emotion transitions as well as their partner's transitions. We used this task to measure how accurate participants were at emotion predictions, how similar they were to their target, and how typical they were as a member of the university community. Participants then completed questionnaires about four

aspects of their personality: values, attitudes, recreational activities, and feelings towards different social groups. We used these personality questionnaires to assess their dyadic similarity and typicality outside the domain of emotion dynamics. Finally, participants completed measures of their relationship and general social well-being. These measures were administered as a component of the larger longitudinal projects and were not included in any current hypotheses or analyses.

Emotion Transition Task. participants made a series of predictions about how a target's emotions might change over time. Specifically, participants judged on a continuous scale from 0% to 100% how likely it was that one mental state (e.g., happiness) would transition into another (e.g., elation) for a particular target.

Participants completed the task for three targets: (i) Self: a participant's ratings of their own emotion transitions, (ii) Target: a participant's ratings of their partner's emotion transitions, and (iii) Group: a participant's ratings of the emotion transitions of the average student at their university. The order of targets was randomized across participants. For the purpose of the present study, only Self ratings and Target ratings were analyzed.

We used the emotion transition ratings to assess four constructs of central interest: accuracy, egocentrism, similarity, and typicality. Accuracy refers to how well one's predictions about their target match the target's actual emotion transitions. To assess a participant's accuracy, we computed the Pearson's correlation between the participant's Target ratings and their target's Self ratings. Egocentrism refers to how much one anchors their social predictions on self-knowledge. To assess how much a participant engaged in egocentric strategies, we computed the Pearson's correlation between the participant's Target ratings and their own Self ratings. Similarity refers to how much two people's emotion transitions corre-

spond to each other. To measure dyadic similarity within the specific domain of emotion transitions, we computed the Pearson's correlation between a participant's Self ratings and their target's Self ratings. Finally, typicality refers how similar one is to the average person in their social environment. To assess a participant's typicality within the specific domain of emotion transitions, we computed the Pearson's correlation between the participant's Self ratings and the average Self ratings of the sample. We applied Fisher's z-transformation to each of these four, respective, correlation indices.

Personality Measures. Participants completed a series of survey questions that measured their personality in four different domains: attitudes, values, recreational activities, and feelings towards different social groups. To measure social and political attitudes, participants indicated, on a seven-point Likert scale, how much they agreed or disagreed with 11 statements (Bahns, Springer, & The, 2015; Bahns et al., 2017). To measure values, participants indicated, on a seven-point Likert scale, how important 12 values (e.g., independence) were to them as guiding principles (Murray, Holmes, Bellavia, Griffin, & Dolderman, 2002). To measure recreational activities, participants indicated, on a five-point Likert scale, how frequently they engaged in 15 different activities (e.g., going to the movies, going out to eat). To measure feelings towards different social groups, participants indicated, on a seven-point Likert scale, how warm or cold they felt towards 10 social groups (e.g., children, women).

Participants' responses on the personality measures were used to measure their domain general dyadic similarity and typicality. To assess domain general dyadic similarity, we computed the Euclidean distance (total absolute difference) between a participant's responses and those of their targets across all personality measures. To assess domain general typical-

ity, we computed the Euclidean distance between a participant's responses and the sample average responses across all personality measures. These domain general scores were then reverse coded, such that higher value means more similar or typical, to facilitate interpretation.

2.1.2 Analyses

In the current study, we aimed to demonstrate the rationality of using the self as a reference point for social predictions. To do so, we conducted three sets of interrelated analyses: (i) egocentrism leads to accurate social predictions: this analysis tested the relation between egocentrism and accuracy (ii) typicality cues people to be egocentric: this analysis tested the relation between participant's typicality and egocentrism (iii) typicality is linked to more accurate prediction at zero acquaintance: this analysis tested the relation between participant typicality and accuracy, as well as whether this relation is mediated by egocentrism and similarity.

For each analysis, we computed the bivariate correlation between variables at the individual level, and then used these scores for correlation and regression analyses across individuals at the sample level. For example, to examine the relation between egocentrism and accuracy, we first computed each participants' egocentrism and accuracy scores. We then calculated the correlation between the two scores across participants. One issue this approach faced was that the individual level correlation scores of different constructs can share underlying covariates. Egocentrism scores were calculated by correlating each participant's Self ratings with their Target ratings, whereas accuracy scores were calculated by correlation each participant's Target ratings with their target's Self ratings. The two con-

struct scores shared the participants' Target ratings as a covariate. Such covariate sharing can obscure the true relation between correlation-based construct scores.

As such, our analyses applied an iterative subset-and-bootstrap procedure to eliminate the confounds produced by covariate sharing. In each iteration of this procedure, we first divided the 75 state pairs from the emotion transition task into two equally sized subsets. Each subset was used to estimate one individual level score. For example, to estimate the relation between egocentrism and accuracy, we first divided the 75 items into two sets (a 37-38 split in a single iteration, but equally sized in the long run). For each participant, we obtained their egocentrism scores by correlating their Self ratings and Target ratings on item set 1, and then applying Fisher's z-transformation. We obtained accuracy scores for the same participant by correlating their Self ratings and their target's Self ratings on item set 2, and then applying Fisher's z-transformation. Once these scores were computed, we bootstrapped over participants by repeatedly resampling, with replacement, the same number of participants as the actual sample. For each sample, we calculated the correlation between egocentrism an accuracy scores. These two steps were repeated in a nested fashion to obtain an empirical distribution of the relation of interest. Simulations showed that such a subset-and-bootstrap procedure can effectively address the covariate-sharing issue, as well as recover true sample-level correlations (see B for details). All analyses that involved covariate-sharing were conducted using this procedure.

The Accuracy of Egocentrism. We examined whether egocentrism constituted an effective strategy for people to achieve accurate emotion predictions. To do so, we tested the correlation between participants' egocentrism scores and accuracy scores. If anchoring social predictions on self-knowledge indeed aids greater accuracy, we expect that egocen-

trism and accuracy scores would be positively correlated.

Rational Cues for Egocentrism. Our main analysis examined whether egocentrism might be informed by perceiver's typicality. To do so, we calculated the correlation between egocentrism scores and perceiver typicality scores. If typicality informs the extent to which people anchor their social predictions on self-knowledge, we should see positive correlations between egocentrism and typicality.

We also attempted to replicate previous work showing that egocentrism is tuned to dyadic similarity. To do so, we calculated the correlation between egocentrism and dyadic similarity scores. If dyadic similarity informs the extent to which people employ egocentric strategies, we expected that egocentrism and dyadic similarity should be positively correlated.

Given that dyadic similarity and dyad members' typicality were inherently correlated, we also examined whether typicality exerted unique influences on people's egocentric strategy. We conducted linear regression analyses, using both similarity scores and perceiver typicality scores to predict egocentrism scores. If typicality accounts for egocentrism above and beyond dyadic similarity, we would expect typicality scores to positively predict egocentrism scores after controlling for similarity scores.

The Accuracy of Rational Egocentrism. We examined whether typicality shaped how accurately people can make emotion predictions. To do so, we calculated the correlation between accuracy scores and perceiver typicality scores. If typicality shapes how accurately one can predict others' emotion transitions at zero acquaintance, we should see positive correlations both between accuracy and perceiver typicality.

Finally, we examined a pathway through which typicality is linked to accuracy. Specifically, we tested the possibility that more typical perceivers achieve greater accuracy because they are more egocentric. In a mediation analysis, we tested whether egocentrism scores accounted for the association between perceiver typicality scores and accuracy scores. If more typical individuals achieve greater accuracy by virtue of relying more on self-knowledge, we expect that egocentrism should mediate the typicality-accuracy link. To ensure the generalizability of this mediation analysis, we conducted a convergent analysis using domain general measures of typicality calculated from the personality measures.

2.1.3 Results

The Accuracy of Egocentrism. In our first analysis, we tested whether egocentrism is a rational strategy for social inferences by measuring the relation between egocentrism and accuracy scores. Egocentrism scores were positively associated with accuracy scores ($r(114) = .50, 95\%CI = [.33, .64]$; Figure 2.1), suggesting that those who anchored their emotion predictions more heavily on self-knowledge also achieved overall greater levels of accuracy about their target. That is, egocentrism was a sound strategy that supported accurate social inferences.

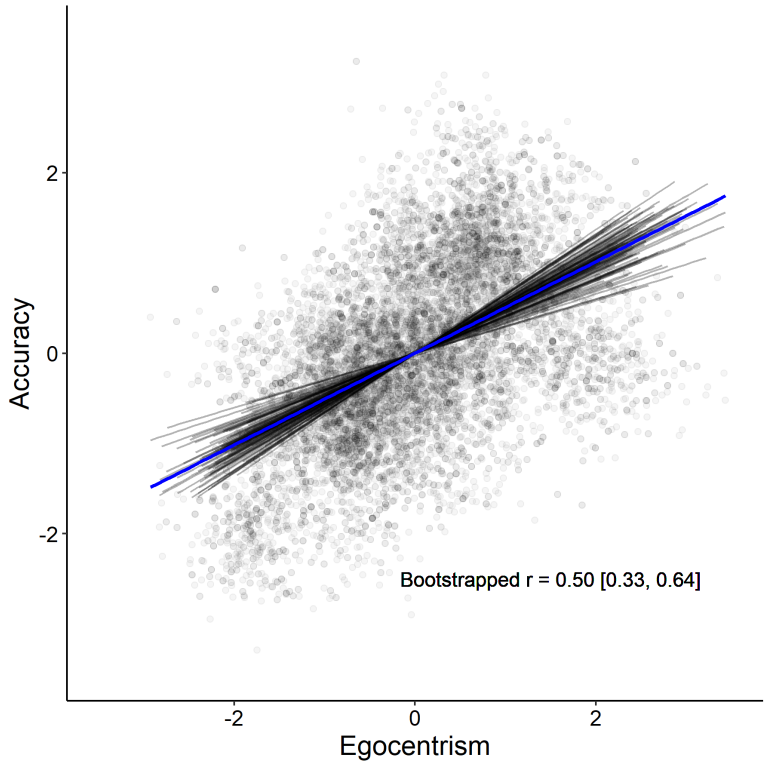

Figure 2.1: Correlation between Egocentrism and Accuracy scores. People who were more egocentric in their emotion predictions also tended to be more accurate. Points represent 100 randomly selected, normalized samples from the bootstrap analysis. Each black line represents the best fitting regression line for a sample. Blue line represents the mean bootstrapped correlation.

Rational Cues for Egocentrism. In our second set of analyses, we examined whether people strategically tune their level of egocentrism based on how typical they are of their social surroundings. That is, we examined whether people are more egocentric when it is rational to do so because they are generally more similar to most people. To do so, we tested the correlation between participants' egocentrism scores and typicality scores (Figure 2.2). We found that typicality positively predicted egocentrism ($r(114) = .60, 95\% CI = [.45, .72]$). We repeated this analyses with domain general typicality, and found the two to

be positively correlated ($r(114) = .28, 95\%CI = [.08, .43]$). These results suggest that participants anchored their predictions more heavily on self-knowledge when they were generally more similar to others in their social environment.

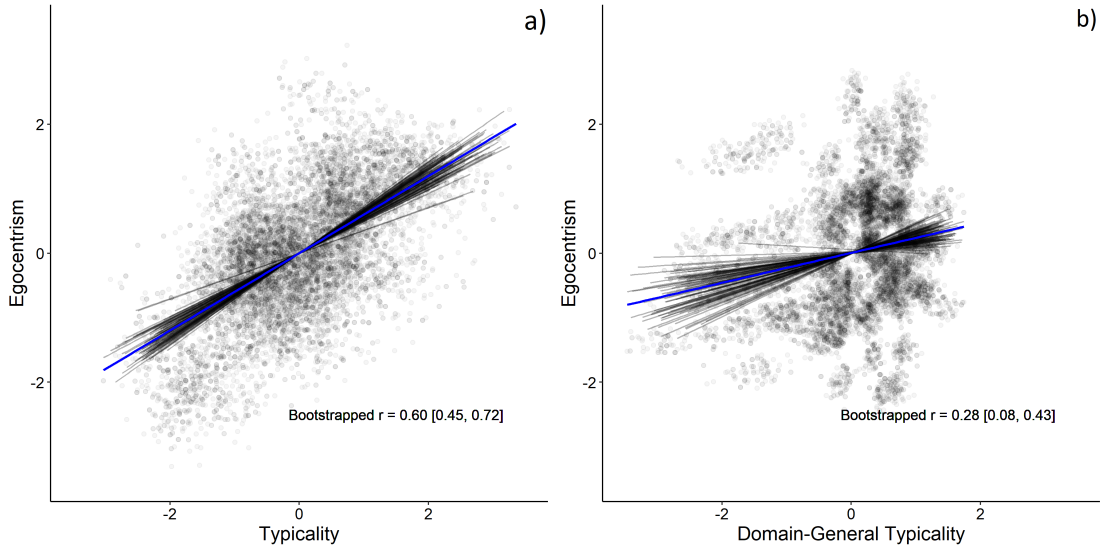

Figure 2.2: Correlation between Egocentrism and Typicality scores. People who were more typical, both in terms of their (a) emotion dynamics and (b) personality, made more egocentric emotion predictions. Points represent 100 bootstrap samples. Each black line represents the best fitting regression line for a sample. Blue line represents the mean bootstrapped correlation.

In addition, we examined whether people might strategically tune their level of egocentrism based on dyadic similarity. That is, we examined whether people are more egocentric when it is rational to do so because they were more similar to their target. To do so, we tested the correlation between participants' egocentrism scores and dyadic similarity scores. We found that similarity positively predicted egocentrism (r(114) = .42, 95% CI = [.25, .58]). We repeated this analyses with domain general similarity and found it to correlate positively with egocentrism (r(114) = .21, 95% CI = [.004, .38]). These results replicate

prior work showing that participants anchored their emotion predictions more heavily on self-knowledge when the target was more similar to themselves.

Both similarity and typicality seem to serve as cues to people to be egocentric in their social inferences about a novel target. We next tested the hypothesis that typicality should serve as a stronger cue than dyadic similarity. In a multiple regression, we used perceiver typicality scores and dyadic similarity scores to simultaneously predict egocentrism scores. As predicted, we found that typicality uniquely predicted egocentrism after controlling for similarity ($b = .47, 95\% CI = [.26, .66]$), and that similarity did not have a unique effect ($b = .16, 95\% CI = [-.07, .38]$). We repeated this multiple regression using domain general similarity and typicality scores. Consistent with the domain-specific analyses, we found a unique effect of typicality after controlling for similarity ($b = .25, 95\% CI = [.01, .49]$), and no a unique effect for similarity controlling for typicality ($b = .04, 95\% CI = [-.20, .28]$). These findings suggested that egocentric strategies were more strongly cued by typicality than by one's similarity to a specific target.

Taken together, these results supported our hypothesis that people adapt their predictive strategies to contextual cues such as dyadic similarity and typicality, and that typicality might shape predictive strategies to a greater extent than similarity when people make predictions about unfamiliar targets.

The Accuracy of Rational Egocentrism. In our third set of analyses, we examined whether these cues to be egocentric led typical perceivers to be more accurate in their inference. We started by investigating whether a perceiver's typicality shaped how accurately they could predict a novel target's emotion transitions. To do so, we tested the correlation between participants' accuracy scores and typicality scores Figure(2.3). We found

that typicality positively predicted accuracy ($r(114) = .43, 95\%CI = [.25, .58]$), suggesting that the more typical a participant was of their social environment, the more accurately they could predict a novel target's emotion transitions. We repeated this analysis with domain general measures of typicality, which were computed using the personality measures. We again found that typicality was positively associated with how accurately participants could predict emotion transitions ($r(114) = .24, 95\%CI = [.03, .42]$). Taken together, these results support the hypothesis that one's ability to make accurate emotion predictions in the absence of detailed knowledge about their target was linked to their own typicality within a social context.

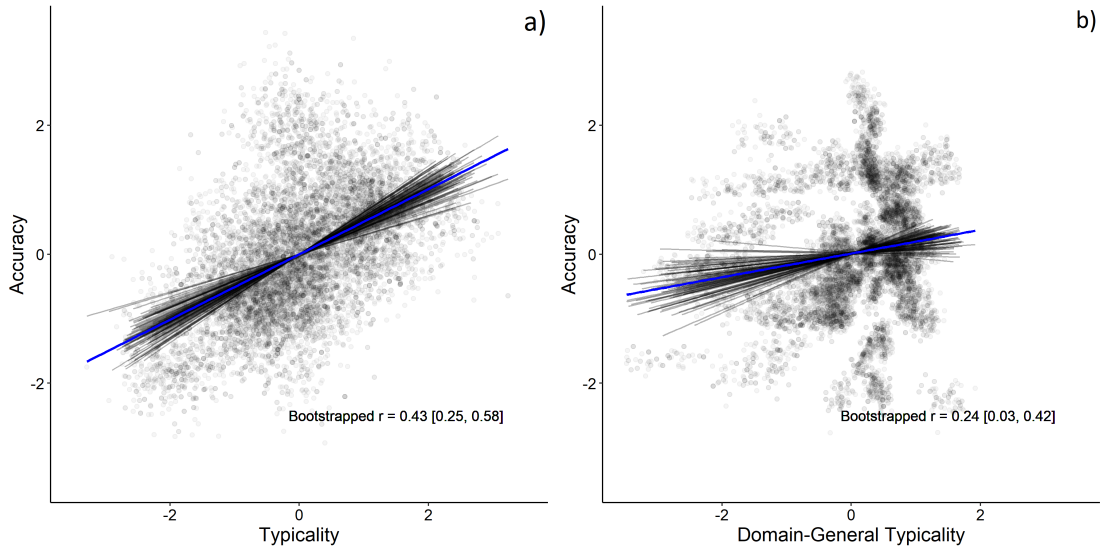

Figure 2.3: Correlation between Accuracy and both (a) domain-specific and (b) domain-general Typicality scores. People who were more typical, both in terms of their emotion dynamics and their personality, predicted novel targets' emotion transitions more accurately. Points represent 100 bootstrap samples. Each black line represents the best fitting regression line for a sample. Blue line represents the mean bootstrapped correlation.

Next, we aimed to explain why more typical people could make more accurate emotion

predictions about novel individuals. To do so, we used mediation models to test two potential pathways. Our hypothesis is that one's typicality leads people to strategically use self-knowledge during social inferences, and that this strategy, in turn, leads to accuracy (Figure 2.4). In a mediation model, we tested whether egocentrism mediated the relation between typicality and accuracy. The total effect using typicality to predict accuracy was $b = .44\,(95\%CI = [.27, .59])$. We detected a positive indirect effect through egocentrism ($b = .21, 95\%CI = [.07, .37]$). The effect of typicality was reduced after mediation, leaving a direct effect of $b = .23\,(95\%CI = [-.01, .46])$ and indicating that the association between typicality and accuracy was mediated through egocentrism. This pattern of results generalized to domain general analyses, where the total effect of domain general typicality in predicting accuracy was b = .22 ($95\%CI = [.01, .41]$). We detected a positive indirect effect through egocentrism ($b = .12, 95\%CI = [.02, .21]$). The effect of domain general typicality was reduced after mediation, leaving a direct effect of $b = .11$ ($95\%CI = [-.08, .27]$), suggesting that egocentrism mediated the association between domain general typicality and accuracy. These results further support the idea that the more typical a person is in general the more they use self-knowledge to make emotion predictions, and thus the more accurate they were.

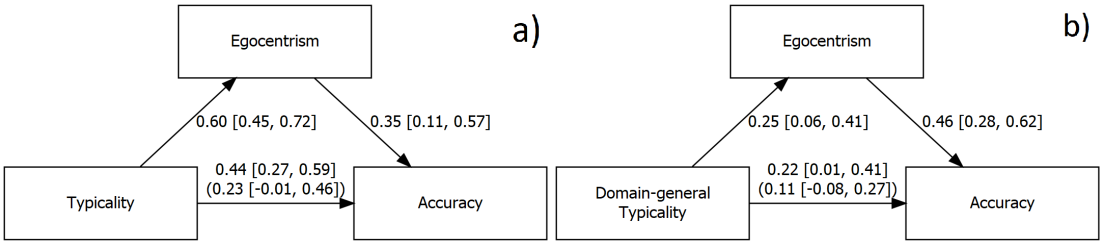

Figure 2.4: Egocentric inference mediated the Typicality-Accuracy link. Typical perceivers more accurately predicted novel targets' emotion transitions, and they did so by virtue of being more egocentric.

Typicality can also be associated with accurate emotion predictions through other pathways, such as interpersonal similarity. Typical individuals are more likely to be similar to any given person in their social context. When people make similarly egocentric predictions, those who are more similar to their specific target will be more accurate. In the Supplemental materials, we include an analysis of whether interpersonal similarity accounts for the link between typicality and accuracy *in addition* to egocentrism.

2.1.4 Discussion

Consistent with our hypotheses, results from this study revealed that an individual's typicality can confer substantial advantages at the outset of new social relationships. More typical individuals engage in more egocentric predictions, which in turn makes it more likely that they make accurate inferences about their targets.

2.2 Study 4. Replicating Study 3

In Study 3, we set out to investigate a set of three interrelated ideas. We showed that egocentrism –using knowledge about the self to make inferences about others –could lead to

accurate predictions about unfamiliar targets. We then showed that people alter the extent to which they employ an egocentric strategy based on similarity cues, both dyadic similarity and their own global typicality. Finally, we showed that more typical individuals make more accurate prediction about unfamiliar others because they have learned to employ egocentric strategies to a greater extent. The analyses of Study 3 were largely exploratory. In the current study, we aimed to conduct a preregistered replication of these findings in an independent sample, collected using a very similar paradigm.

2.2.1 Methods

Participants. Participants in Study 4 were first-year undergraduate students at Princeton University. Each student signed up for the study with one roommate at the outset of the fall semester. Participants were recruited as the first wave of a second longitudinal study aimed to replicate and extend findings from Study 3. Subsequent waves were not analyzed for the current work. The same sample size of 40 was set a priori and 58 newly formed roommate dyads were recruited (N = 116, 66 women; Mean age = 18.16, SD = 0.88; 41.4% White, 37.9% Asian, 12.1% more than one ethnicity, 4.3% others, 3.4% Black or African American, .9% Native Hawaiian and Pacific Islander; 18% Hispanic or Latinx). All roommates were assigned by university housing rather than self-selected, and had known each other for 27.5 days on average at the time of study completion (range = 8-81 days). Two dyads were excluded prior to analyses due to prior acquaintanceship (relationship durations = 140 days and 176 days). All participants provided informed consent in a manner approved by the Princeton University Institutional Review Board.

Procedures. The measures and procedures used in Study 4 were identical to those in Study 3. Participants completed the emotion prediction task. Their responses were used to compute domain-specific scores of egocentrism, similarity, typicality, and accuracy. Participants' responses personality measures were again used to compute domain-general versions of those scores. These data were used to conduct a preregistered replication of the three main sets of hypotheses from Study 3: (i) more egocentric emotion predictions will be more accurate, (ii) people who are more typical should make more egocentric predictions, (iii) people who are more typical should make more accurate predictions because they tend to be more egocentric.

2.2.2 Results.

The Accuracy of Egocentrism. In our first analysis we tested whether egocentrism is a rational strategy, in that it supports accurate social inferences. Results replicated the positive association between participants' egocentrism scores and accuracy scores ($r(110) = .50, 95\%CI = [.31, .65]$; Figure 2.5). These results further supported the hypothesis that egocentrism can be an effective strategy for social prediction.

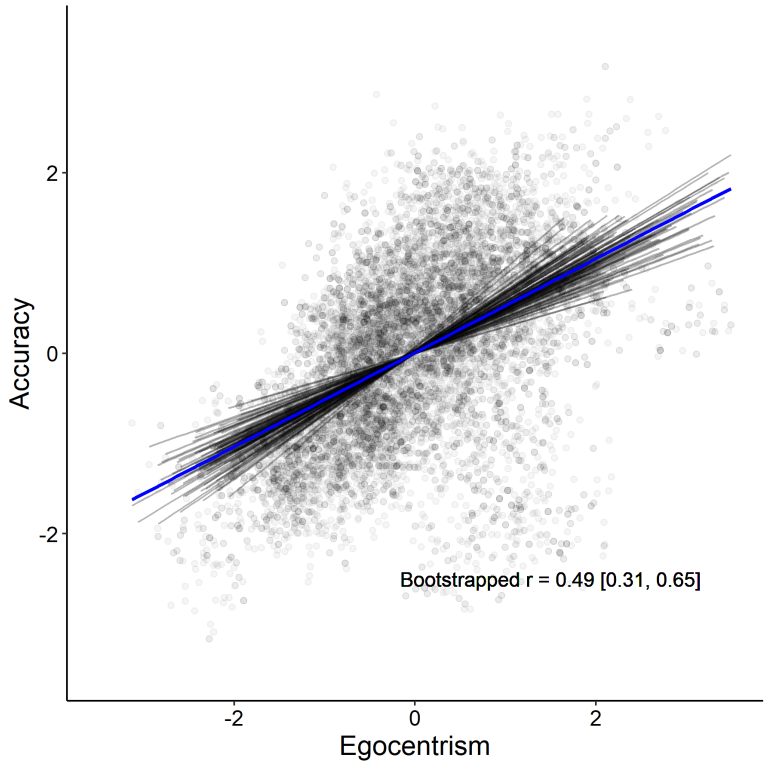

Figure 2.5: Correlation between Egocentrism and Accuracy scores in the Replication sample. More egocentric perceivers made more accurate emotion predictions. Points represent 100 bootstrap samples. Each black line represents the best fitting regression line for a sample. Blue line represents the mean bootstrapped correlation.

Rational Cues for Egocentrism. Our second set of analyses tested whether people adjust the use of egocentric strategies based on how typical they are of their social environment. To this end, we tested the correlation between participants' egocentrism scores and typicality scores. The results again fully replicated findings from Study 3 (Figure 2.6). We found a positive association between typicality scores and egocentrism ($r(110) = .71, 95\%CI = [.58, .81]$). Likewise, domain general typicality positively predicted egocentrism ($r(110) = .41, 95\%CI = [.26, .54]$).

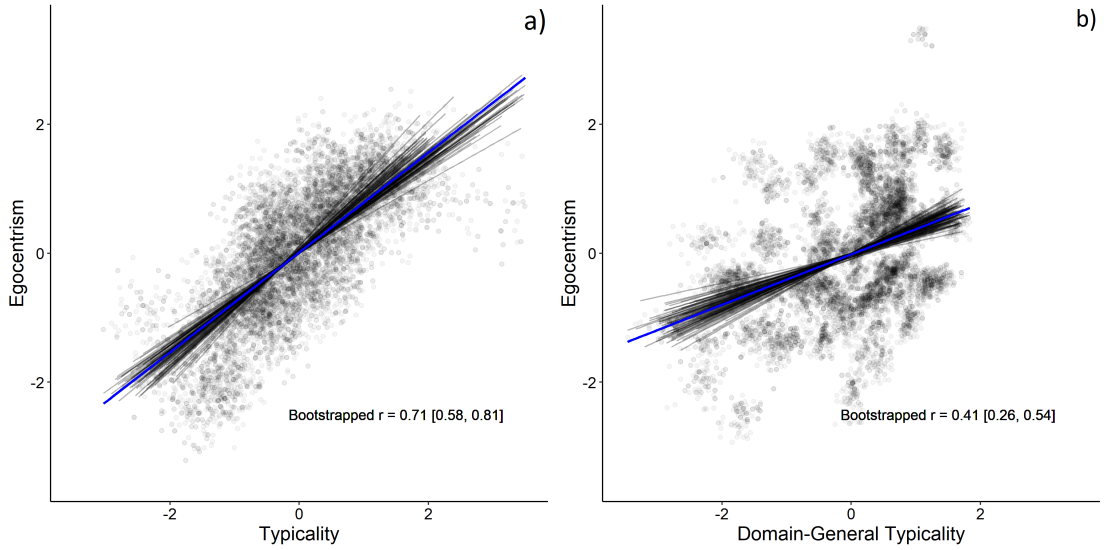

Figure 2.6: Correlation between Egocentrism and both (a) domain-specific and (b) domain-general Typicality scores from the replication sample. People who were more typical relied more heavily on their self-knowledge to predict others' emotions. Points represent 100 bootstrap samples. Each black line represents the best fitting regression line for a sample. Blue line represents the mean bootstrapped correlation.

We also tested whether egocentric strategies were informed by how similar people were to their specific partner. To do so, we tested the correlation between participants' egocentrism scores and similarity scores. We replicated the positive association between dyadic similarity and egocentrism ($r(110) = .46, 95\%CI = [.28, .62]$). This association extended to domain general similarity, which was also positively associated with egocentrism ($r(110) = .28, 95\%CI = [.09, .44]$). These results show that people tailor their inference strategy based on how relevant self-knowledge is likely to be.

Importantly, we again tested the hypothesis that typicality should more strongly cue egocentrism than dyadic similarity. In a multiple regression, we found that typicality uniquely predicted egocentrism after controlling for similarity ($b = .60, 95\%CI = [.41, .77]$), but

similarity did not have a unique effect after controlling for typicality ($b = .11, 95\%CI = [-.08, .31]$). When the regression analysis was repeated using domain general measures, typicality again uniquely predicted egocentrism controlling for similarity ($b = .36, 95\%CI = [.17, .54]$), and similarity did not uniquely predict egocentrism after controlling for typicality ($b = .14, 95\%CI = [-.05, .32]$). These findings suggested that egocentric strategies were more strongly cued by one's typicality than by one's similarity to a specific target. These results corroborated Study 3 finding that typicality was a strong cue for egocentric strategies than one's similarity to a specific novel target.

Taken together, this set of findings further supported the idea that people adapt how much they rely on self-knowledge when making predictions about novel individuals. This adjustment tracks more closely with the perceiver's typicality in the social environment than with their similarity to the specific target.

The Accuracy of Rational Egocentrism. In the third set of analyses, we examined whether typical perceivers were more accurate in their emotion prediction. We replicated Study 3 findings on the association between typicality and accuracy. We found that a perceiver's typicality was associated with how accurately they predicted a novel individual's emotion transitions (Figure 2.7). Typicality scores were positively correlated with accuracy scores ($r(110) = .46, 95\%CI = [.28, .61]$). Furthermore, the typicality-accuracy association extended to domain general analyses, where domain general typicality was also positively correlated with accuracy ($r(110) = .37, 95\%CI = [.21, .52]$). Together, these results lent further credence to the idea that more typical individuals could make more accurate emotion prediction in the absence of knowledge about a specific target.

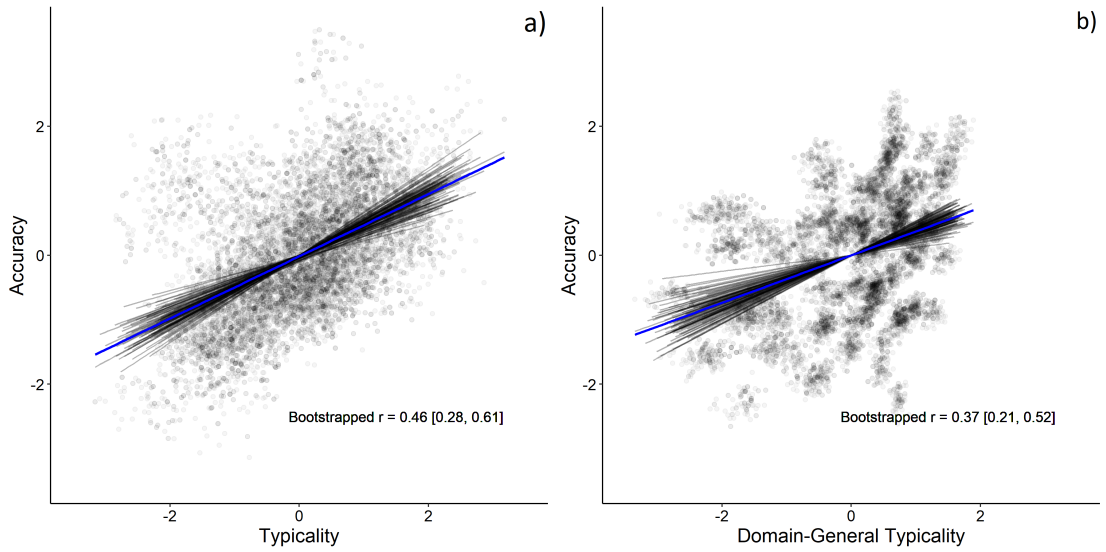

Figure 2.7: Correlation between Accuracy and both (a) domain-specific and (b) domain-general Typicality scores in the replication sample. Typical perceivers more accurately predicted novel targets' emotion transitions. Points represent 100 bootstrap samples. Each black line represents the best fitting regression line for a sample. Blue line represents the mean bootstrapped correlation.

We then tested the hypothesized pathway that might account for the typicality-accuracy association. More specifically, we replicated Study 3 results showing that typicality influenced accuracy by shaping a person's strategic use of self-knowledge (Figure 2.8). In a mediation model, egocentrism score partially mediated the association between typicality and accuracy. The total effect of typicality predicting accuracy was $b = .46$ ($95\% CI = [.27, .64]$). We detected a positive indirect effect through similarity ($b = .23, 95\% CI = [.06, .45]$). After mediation, the effect of typicality was reduced, leaving a direct effect of typicality of $b = .23$ ($95\% CI = [-.05, .48]$). Likewise, the mediating effect of egocentrism held in domain general analyses. Egocentrism scores mediated the association between domain general typicality and accuracy. The total effect of domain general typicality predict-

ing accuracy was .33 ($95\%CI = [.15, .49]$). We detected a positive indirect effect through egocentrism ($b = .18, 95\%CI = [.07, .30]$). The effect of typicality was reduced to $b = .15$ ($95\%CI = [-.03, .35]$) after mediation. These replications strengthened the conclusion that being more typical of one's environment can aid individuals to make more accurate predictions about others in the absence of person knowledge. They also further supported the idea that more typical individuals achieve more accurate predictions by strategically relying more on self-knowledge to make predictions. Replication analyses on the additional mediating effect of interpersonal similarity can be found in the Supplement.

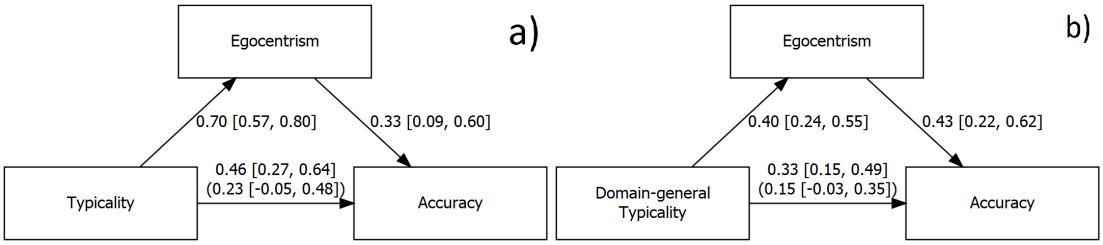

Figure 2.8: Egocentric inference mediated the Typicality-Accuracy link in the replication sample. Typical perceivers more accurately predicted novel targets' emotion transitions, and they did so by virtue of being more egocentric.

2.3 Discussion

To create and maintain meaningful interpersonal connections, it is critical that people anticipate others' thoughts and emotions. Imagine meeting up with someone who has recently gone through some stressful events and was quite distraught. How you approach the interaction will depend on how you expect them to be feeling. If you expect them to still be affected, you might try to be a lot more sensitive than if you expect that they have already gone back to their usual self. If this person is an old friend, you might be able to form

your expectations based on the kind of person they are or how they've dealt with similar situations in the past. However, what if this is someone you've only recently met and had only very limited personal knowledge about? This task of social prediction can be highly challenging, especially in the context of meeting new people, where the perceiver might operate in a vacuum of knowledge about the target. To solve this challenge, people routinely resort to an egocentric strategy, anchoring their social prediction on what they know about themselves. Here we show that egocentrism provides an effective, rational solution to the challenge of social prediction. Across an initial study and a pre-registered replication study, we found robust evidence that egocentric emotion prediction tended to be accurate. Further, we see that people employ this egocentric strategy when doing so was particularly likely to lead to accurate predictions, such as when they were more typical members of their community.

We found that egocentrism can be an overall effective, rational strategy for social prediction. People who anchored their emotion prediction more on self-knowledge also more accurately predicted an unfamiliar target's emotion transitions. This finding aligns with previous research demonstrating the potential rationality of egocentrism in other domains of social judgment, such as attitudes, beliefs, and preferences (Hoch, 1987; Woo & Mitchell, 2020). While people often consider egocentrism a miscalibrated heuristic for thinking about other people, here we show that egocentrism, instead, can be a rational strategy that supports accurate social inferences. Bias is relative. If people have unlimited processing power and all the relevant information about their social target, it would seem ill-advised to make predictions based only on what they know about themselves. However, such an ideal scenario is improbable in the social world, especially during early encounters with a novel

social target. Social predictions about a novel target would be a complete shot in the dark if perceivers were not able to draw upon their own self-knowledge.

We also found that people rationally tune their egocentric strategy, and use self-knowledge to a greater extent when doing so would likely lead to more accurate predictions. More specifically, we found that people's degree of egocentrism tracked social similarity at two levels –dyadic similarity and typicality. People make more egocentric predictions when they are more similar to their target. This dovetails with the extensive literature showing exacerbated egocentric bias for the ingroup (Ames, 2004b; Robbins & Krueger, 2005). However, here we found that the relation between similarity and egocentrism can be accounted for by typicality. When we observed a positive relation between similarity and egocentrism in the current data, it was because both similarity and accuracy were positively associated with typicality. Furthermore, the tendency for typical perceivers to be more egocentric could be a pathway through which they manage to more accurately predict the emotions of novel, unfamiliar targets. On average, typical perceivers are more likely to be similar to any given person they encounter in their social lives. With repeated feedback from such "chronic similarity," those who are typical of their social context learn, over their lifetime, that their self-knowledge serves as a good proxy for others. This drives them to default more to egocentrism and adapt less to any given dyadic context and serve them particularly well in a highly homogenous community.

On the other hand, those who are less typical might learn to adjust their prediction strategy on a much shorter timescale. They might do so by adapting their predictions in a person-specific manner, particularly at the outset of interacting with a new social partner. However, by attempting to tune their predictions to a greater extent even when their

self-knowledge would have sufficed, less typical individuals can end up making less accurate emotion predictions at the outset of interacting with a new social partner. An intriguing possibility is that typicality influences not only how accurately people can make predictions about novel individuals during early encounters, but also how their accuracy will change over the time course of interpersonal relationship development. If more typical perceivers have indeed learned to adapt less to each given specific target, they might fail to learn and incorporate useful, person-specific knowledge about their partner, and thus show less improvement in their prediction accuracy over time. In contrast, less typical perceivers might show greater flexibility in encoding and employing target-specific knowledge in their predictions, and become more accurate over time. Future research should test these predictions by tracking typicality and predictive accuracy over time using longitudinal designs.

Conventionally, rationality has meant acting in accordance with what logic and probability theory dictate. More recently, however, researchers have argued that the rationality of a behavior should be evaluated alongside the ecological contexts in which the decision problem takes place (Marr, 2010). That is, to assess whether people act or make decisions rationally, we must first carefully characterize the computational problem people need to solve, taking into account its natural statistics and the decision maker's resource constraints (Lieder & Griffiths, 2020). The current studies demonstrated that egocentric strategies provide rational solutions in a naturally occurring, ecologically valid context where emotion prediction occurs, namely a university community.

Nonetheless, the rationality of egocentrism is qualified by two characteristics of naturalistic emotion prediction. First, real-life social contexts are often characterized by homophily (McPherson, Smith-Lovin, & Cook, 2001). Through both structural and assortative mech-

anisms, people that are closely linked in a social network tend to be highly similar. When there is a high level of interpersonal similarity on average, people can make more accurate predictions about others simply by being egocentric. In contrast, when people try to integrate other sources of information, they often rely on stereotypical beliefs. Such information could often be unreliable or not rooted in reality, and thus make the predictions less accurate (Eyal et al., 2018). That said, egocentrism might be rational only in a homogenous environment. Both samples tested in the current studies were recruited from a closed university community that has been heavily pre-selected. Our participants might therefore have been overall more similar to each other than a random sample of people, thus artificially inflating the link between egocentric strategies and prediction accuracy. Future research should aim to explore how social network homogeneity impacts people's social prediction strategies and their efficacy.

Second, it is possible that egocentrism confers advantage on accuracy only in specific domains of social predictions. In the current work, we tested how well people can make predictions about emotion transitions. While people can vary widely in their emotion dynamics, there are overarching, central regularities of how emotions tend to change over time (Cunningham et al., 2013; Thornton & Tamir, 2017). Due to these latent regularities, people might be more similar to each other in their emotion transitions than in other domains, rendering egocentrism a particularly effective decision strategy for emotion prediction. In other social domains, however, such high levels of interpersonal overlap might not exist. One such example is political beliefs. When there is a wide range of possible beliefs and ideologies and people are highly polarized, it might not be prudent to assume an unfamiliar target will be like the self. Future research on egocentrism in any social domain

should aim to characterize the extent of organically occurring interpersonal similarity before drawing conclusions about bias and rationality.

To investigate the relations among typicality, egocentrism, and the accuracy of emotion prediction in an ecologically valid manner, the current work employed a correlational approach. While analytical methods such as mediation model and multiple regression can add confidence to the hypothesized causal directions, we ultimately could only draw correlational conclusions. Future research can experimentally manipulate both the homogeneity of the prediction context and the typicality of the perceiver using a paradigm where participants learn about a community of fictional agents. By intervening on the statistical characteristics of the decision problems, we can draw more direct and more fine-grained conclusions about both the efficacy of egocentric prediction, and how people might use such strategies adaptively.

In sum, the current work applies approaches from rational analysis to demonstrate that egocentric strategies are rational, given the natural contexts in which emotion prediction occurs. These results call researchers to reconsider the conventional dichotomy between bias versus rationality. Rather than attempting to classify cognitive phenomena neatly into one of the two categories, future research should consider cognition and its context in a combined, graded manner. Most importantly, these results provided a somewhat paradoxical prescription of how social predictions should be made in a divided world –that sometimes, the best way to step into others' shoes is to try on your own.

3

Latent Social Structure Facilitates Partner-specific Emotion Prediction from Sparse Observations

People's thoughts and feelings are constantly in flux. Someone who is red with rage might, after a few moments, come to experience intense remorse. While there are

systematic patterns underlying the ebbs and flows of emotion states across the population (Thornton & Tamir, 2017), these dynamics also vary widely across individuals. Everyday social interaction depends on the ability to anticipate how others' emotions may change over time, but if everyone has their own idiosyncratic dynamics, predictions based on population averages are likely to fall short. In other words, accurate prediction requires *partner-specific* knowledge. Past research has shown that people who have rich partner-specific knowledge, often in the contexts of close friends or long term acquaintances, can more accurately predict both the contents (Stinson & Ickes, 1992) and the dynamics (Zhao et al., 2020) of their partner's emotions . The better we know someone, the better we can anticipate their ups and downs. But what does it take to acquire such partner-specific knowledge?

3.0.1 Two Types of Learning

One possibility is that partner-specific knowledge is slowly and laboriously accumulated from repeated observations. As people and their partner experience episodes of shared situations, events, and activities, they make observations about how one another's emotions change over time. Observations are counted and tallied to form a mental model of a specific target's emotion transitions. If this were the only type of information and learning at play, we would predict that the accuracy of partner-specific predictions improves gradually over the long time course of relationship development. Contrary to these predictions, however, recent results have found remarkably accurate partner-specific predictions within the first few weeks of a relationship. Zhao et al.(2020) recruited first-year college roommates who had little to no prior knowledge about each other. They were asked to predict each other's

emotion transitions at two time points, once within the first six weeks of the academic year, and once towards the end of the academic year. Surprisingly, the largest increases in accuracy were found within the first few weeks of the initial wave of assessment, while further improvements measured at the end of the year were marginal.

These findings suggest that, rather than relying on bottom-up associative mechanisms that slowly tally observations, partner-specific predictions may be supported by strong top-down induction-like mechanisms. Within such a short time frame, the roommates would be unable to directly observe and estimate the probability of every possible emotion transition they were tested on. In the absence of such direct observations, they must infer the likelihood of unobserved emotion transitions from sparse observations. This type of inductive learning depends on the latent statistical structures of emotion dynamics in the population. While the transition probability between any pair of emotion states can, by itself, be simply described using its central tendency and variance, transition probabilities across different emotion pairs are likely not independent of each other. For example, covariances between different emotion transitions might have clustered structures. That is, there might be different *types* of people who exhibit distinct patterns of emotion transitions. If people indeed make use of structured statistics in emotion prediction, they might be able to extrapolate from a few observed emotion transitions to the unobserved parts of the transitions space. Furthermore, if this inductive knowledge accurately corresponds to the statistical structure of emotion dynamics in the real world, people can become more accurate about specific targets even at very early stages of relationship development.

In the current work, we formalize this hypothesis using probabilistic models and investigate the extent to which individuals may represent the latent social structure underlying

emotion dynamics. Furthermore, we explore whether people can in fact take advantage of these statistical structures and rapidly learn to make accurate, partner specific emotion predictions for novel social targets. More specifically, we explore two possible forms of statistical structures: as discrete social categories or groups and as a continuous space with meaningful dimensions.

3.0.2 Categorical Representations

Research in social cognition has demonstrated people's pervasive tendency of thinking about others in categorical terms. Salient social categories such as gender and race are often automatically encoded (Weisman, Johnson, & Shutts, 2015). These categories guide how people people represent others in their memory (Srull & Wyer, 1989), what people expect others to prefer (Gershman, Pouncy, & Gweon, 2017) and how they expect others, and even themselves (Hogg, 2020), to think and act (Tajfel, 2001).

It is possible that categorical, group-based representations guide people's inductive emotion predictions. More specifically, people might hold the intuitive theory that there are *types* of people that share similar patterns of emotion transitions. That is, a target's observed emotion transitions contain information about what category they might fall into. Once a perceiver has observed a few emotion transitions from a target, they can quickly place the target into a mental category, and then use their existing beliefs about this category to make predictions about other emotion transitions.

For example, one intuitive theory about emotion categories might be that there are two types of people: those who are happy and those who are sad. Happy people tend to both gravitate towards and stay in positive emotions, whereas sad people gravitate towards and

stay in negative emotions. Given these beliefs, when people observe someone transition from a negative state into a positive one, they can quickly *classify* the target as a happy person. From there, they can predict that target is likely to stay in the positive state, and unlikely to transition into a negative state, even though they might not have observed those transitions from the target. From our perspective, however, the underlying categories guiding emotion predictions need not conform to extant categories of emotion or to other common social divisions typically examined in social cognition.

3.0.3 Dimensional Representations

Dimensional representation is the idea that complex system of observations, knowledge, and concepts can be explained as a combination of a smaller number of dimensions. Such an idea has be featured prominently in social psychology. For example, the stereotype content model (Cuddy, Fiske, & Glick, 2008) suggests that social groups are mentally organized by two dimensions: warmth and competence. While the stereotype content model is primarily applied to understand mental representations of entire collectives, specific individuals may also be embedded in the same dimensional space. People intuitively interpret others' behaviors in terms of having more or less of a given personality trait, and modern theories about personality suggest that they can also be accounted for using a low-dimensional model that contains only five or six dimensions (John, Donahue, & Kentle, 1991; Saucier, 2009).

Representations of emotions and emotion transitions may also be organized in a low-dimensional space. For example, the circumplex model (J. A. Russell & Barrett, 1999) suggests that affect can be described using two orthogonal dimensions: valence and arousal.

More recent work has proposed an extended three-dimensional space defining mental states in terms of rationality, social impact, and valance, which accounts for both behavioral ratings and neural representations of mental state concepts (Tamir, Thornton, Contreras, & Mitchell, 2016; Tamir & Thornton, 2018). Importantly, these dimensions accurately predict transition probabilities. That is, emotion transitions involving states that are nearby in this latent space tend to have similar transition probabilities to other states, suggesting that the 3-D model might be useful for generalization to novel emotions. If a dimensional structure does underlie people's understanding of emotion dynamics, such an organization should also guide inductive inferences. Once a perceiver has observed a handful of emotion transitions from a specific social partner, they should be able to infer where that partner is likely to be situated in their latent dimensional space, and therefore make more accurate predictions about unobserved transitions.

3.0.4 The Current Studies

Using a combination of computational and behavioral methods, the current research aims to achieve two main goals. First, we aim to uncover the latent structures in how emotion dynamics vary across individuals in the population. To do so, we instantiate probabilistic models with categorical or dimensional structures, in order to see if either form can capture the statistical regularities in ground truth emotion transition data. Second, we aim to determine whether people draw upon these regularities to learn inductively about others' emotion dynamics. We use the trained models as "ideal observers" to represent the structures available to social observers across their lifetime, and then generate model predictions about how people *should* make partner-specific predictions based on only a single example

of an emotion transitions. Finally, we evaluate human judgments in the same one-shot prediction task against model prescriptions to ascertain how well the structures learned by the models corresponded to the structured knowledge that enable people to rapidly individuate their understanding of emotion dynamics.

3.1 Study 5. Modeling Latent Social Structures of Emotion Dynamics

In this section, we use an existing dataset to examine the *actual* statistical structure of emotion transitions across the population. This ground-truth social structure provides a natural starting point for three reasons. First, if there is evidence of real social covariation between emotion transitions, we expect that it should be reasonably captured using state-of-the-art statistical methods. Second, if we *do* discover ground-truth structure in the population, it provides a "rational" proxy for what an individual may be able to learn over the course of a lifetime observing different social partners. Our dataset can thus serve as training data, allowing a cognitive model to explain how this structure can be used to make inferences from sparse data in subsequent experiments. Third, large existing datasets provide a benchmark for comparing different idealized representations of social knowledge; to the extent that one representation explains the ground-truth structure better than another, we may expect this model to also serve as a more adaptive basis for human inferences.

3.1.1 Dataset

Our modeling target is based on a dataset of self-reported emotion transitions from previous research. The dataset contained three independent samples of college aged participants. A total of 353 participants were asked to judge their own likelihood of one emotion state

(e.g. happiness) transitioning into another state (e.g. elation). On each trial, participants saw a pair of emotional states (e.g., happiness → elation). Participants used a continuous scale from 0% to 100% to rate the probability that, if they themselves were currently experiencing the first state, they would next experience the second state. The states used in this task consisted of three independent sets of five emotional states and participants provided judgements for all possible pairs of states within each set for a total of 75 trials (Thornton & Tamir, 2017).

Our pre-processing pipeline consisted of two steps. We first excluded 15 auto-transitions (e.g happiness → happiness), which were consistently rated highly across participants and had little variance across individuals. We then converted the original integer scale yielded by the slider response variable, $[0, 100]$, to the unit interval $[0, 1]$. Because the endpoints of this interval are not inside the support of standard probability distributions over the interval (e.g. the beta distribution), we padded the end points: zeros were changed to 0.01, and ones were changed to 0.99. Thus, our final dataset can be viewed as an $N \times 60$ matrix where each row represents a single participant's responses and each column represents a particular transition.

3.1.2 Computational models of social structure learning

In this section we will formally specify several models of learning social structure in this dataset and discuss each model's basic properties. We begin by introducing a simple, unstructured baseline model. The baseline model assumes that the emotion transitions are independent of each other and only learns the overall prevalence and variability of each transition in the population. We then compare this baseline to two more sophisticated

models of structured social knowledge, roughly allowing for co-variation within categorical or dimensional representations, respectively.

Independent attribute baseline

The simplest model of social structure is an assumption that everyone varies independently around some central tendency. For example, we may learn that people statistically tend to be *sad* after they are *frustrated* (i.e. the population *mean*) and acknowledge that some individuals are more or less likely to experience this transition than others (i.e. the population *variation*). Importantly, however, this model assumes that different attributes are independent – learning that an individual is more likely to experience one transition does not provide any information about the other transitions. Formally, we can write this baseline model as a statistical inference problem assuming independent distributions for each transition attribute. Because each observation lies in the unit interval, we model the data using a *beta* distribution. This distribution has two parameters, $a > 0$ and $b > 0$, which are able to capture a range of distributions over the interval.

This model assumes that all responses for a particular transition i are generated by a beta distribution with parameters (a_i, b_i). The statistical problem is to infer what parameter values best explain the data. This can be written as a Bayesian inference problem:

$$P(a_i, b_i|D) \propto P(D|a_i, b_i)P(a_i, b_i)$$

where $P(a_i, b_i|D)$ is the *posterior* distribution representing our beliefs about the parameters after observing responses from many participants. According to Bayes rule, the posterior is derived by re-normalizing the product of two terms: a *likelihood* $P(D|a_i, b_i)$ representing

how well a beta distribution $Beta(a_i, b_i)$ explains the data, and a *prior* $P(a_i, b_i)$ representing the background probability of different parameter values, which we assume is given by: $a_i, b_i \sim Gamma(2, 1/3)$. In practice, we fit this model using maximum likelihood variational inference, searching for the single values of a_i, b_i that best explain the data.

Social groups

While the independent attribute baseline model is able to learn overall trends in emotion dynamics and make relatively accurate predictions about whether a new person would experience particular transitions, it does not allow for inductive *generalization* from one transition to another. Because transitions are assumed to be independent, the best guess about a transition is always the same, no matter how much you've learned about a particular person.

We have proposed an alternative hypothesis: that social knowledge may guide structured predictions about the relationships between emotions. For example, we may say that the *kind of person* that can quickly recover from being sad and transition to a happier state may also be the kind of person that quickly transitions from being elated to being frustrated. The idea that we represent categorical archetypes, such as a *volatile* kind of person or a *steady* kind of person and rapidly assign new social partners into these archetypes requires a different kind of model. Formally, this idea is captured by a latent *mixture* model (Gershman et al., 2017). This model assumes there are K different latent groups, and each person i belongs to exactly one group, where their group assignment is given by $z_i \in \{1, \ldots, K\}$.

Under this model, we may not initially know which people belong to which groups, or

even what the groups are, but by observing many different individuals, we discover that some people are more similar. The key idea is that particular transition distributions, given by a particular set of a_{z_i}, b_{z_i} parameter, are *shared* across members of a group, and these distributions can be jointly inferred along with the membership of each group.

$$\begin{aligned} P(a_{z_i}, b_{z_i}, z_i|D) &\propto P(D|a_{z_i}, b_{z_i}, z_i)P(a_{z_i}, b_{z_i}, z_i) \\ &= P(D|a_{z_i}, b_{z_i})P(a_{z_i}, b_{z_i}|z_i)P(z_i) \end{aligned}$$

Critically, after learning some distinct group properties, $\{z\}$, this model suggests that we may use background knowledge of what different *groups* are like to predict the membership z^* of a new partner based only on a small number of observations D^* from that partner: $P(z * |D^*)$. This group inference then allows for more well-calibrated predictions about other transitions, because we are using the expected *group* distribution rather than the population distribution.

Latent dimensional structure

The categorical, or group-based, model introduced in the previous section aims to discover a small number of discrete 'clusters' that are mixed together in the dataset. While it is more flexible than our *independent attribute baseline*, the social groups model is still rigid in several ways. For one, it assumes that the entire *holistic* matrix of emotion transitions is shared within a group. If some transitions are best explained by membership in one group and other transitions are best explained by membership in another group, there is no way to explain that individual as a *combination* of the two groups: they must be assigned to one or

the other.

In this section, we relax this assumption by allowing *dimensional* representations to guide inductive inference about emotion transitions. Formally, a dimensional representation can be specified by a Latent Dirichlet Allocation model (LDA; Blei, Ng, & Jordan, 2003). This model is commonly used to explain documents, such as New York Times articles, as a mixture of underlying "topics". Rather than assuming a discrete number of clusters and assigning each article to a single cluster, we assume some underlying set of dimensions, or topics, along which articles vary and aim to infer the specific mixture for each article.

In our setting, we assume there is an underlying set of components or topics, z_k, which generate different emotion transition matrices. These components play a similar role to the groups in the previous section: each component represents an independent distribution over emotion transitions with parameters (a_k, b_k). Unlike the previous section, however, we assume that the emotion transitions elicited from individual i are explained by some idiosyncratic combination of the components given by *mixture weights* w_i that sum to one. Our problem is to jointly infer the properties of the underlying components $\mathbf{z} = (a_k, b_k)_{k=1}^K$, and which mixture weights w_i best explain each participant i. In the Bayesian framework, we assume the following priors for these weights:

$$\alpha \sim \text{Gamma}(1/K, 1)$$

$$w_i \sim \text{Dirichlet}(\alpha)$$

$$a_k, b_k \sim \text{Gamma}(2, 1/3)$$

where α is a hyper-prior giving the overall prevalence of the different components w_i in the population. We define our likelihood by marginalizing over the components:

$$\begin{aligned} P(w_i, \mathbf{z}, \alpha|D) &\propto P(D|w_i, \mathbf{z}, \alpha)P(w_i, \mathbf{z}, \alpha) \\ &= \sum_{k=1}^{K} w_i \cdot P(D|a_k, b_k)P(a_k, b_k)P(w_i|\alpha)P(\alpha) \end{aligned}$$

The social group based, categorical model and the latent dimensional model share superficial similarities. Both are mixture models and assume that the observed data are generated by proportionally mixing several latent distributions. However, the two models are distinguished by the level at which the mixing occurs. In the group model, the mixture is the population. Each person is associated with only one of the K underlying components, and making rapid inference about a novel individual is about identifying which component they are associated with. In the dimensional model, on the other hand, the mixture is the person. Each person is associated with *all K* underlying components to varying degrees, and making rapid inference about a novel individual is about estimating the degree of association between a person and each of the components.

It is also worth noting that under $K = 1$, the categorical and the dimensional models are identical to each other, as well as to the independent baseline model.

3.1.3 Model Comparison

We first assessed which model better captured the ground-truth statistical structure of emotion transitions. Can people be divided into groups based on their emotion transitions, or

are the regularities in emotion dynamics best accounted for by a low dimensional space? Furthermore, we aimed to identify the optimal level of model complexity - how many latent groups or dimensions should be included in order to describe the transition data while still remaining parsimonious?

To do so, we fitted both categorical and dimensional models with K ranging from 1 (baseline) to 10 to the same dataset. Model fitting was conducted using the Pyro library (Bingham et al., 2018), which allowed the models to be expressed in a python-based probabilistic programming language, and enabled automatic MAP parameter estimation using variational inference.

To assess how well each model described the data, we computed the sum of log probability of the observed data given the parameter point estimates learned by the models. With a larger number of assumed latent components, the models are increasingly complex, and should be able to better describe the data. Thus, we should expect to see higher total log probability as K increases. Moreover, similar to the logic behind principal components analyses, we expect each additional component to account for a decreasing amount of additional information in the data. The optimal K can thus be determined by plotting marginal likelihood against K, and identifying the inflection point in the graph.

The model fitting results can be seen in 3.1. Contrary to expectation, the total log probability of data did not increase as the models became increasingly complex. For the categorical models, performance decreased linearly as K increased. The baseline model assuming that there is only one group out performed models assuming multiple groups in fitting to the data.

For the dimensional models, model performance held constant for K from 1 to 3, but

decreased linearly after. Consistent with past literature on emotions (Thornton & Tamir, 2017), there's some evidence that the regularities in emotion transitions can be captured with a low dimensional space with 2 to 3 dimensions. However, the 2 or 3 dimensional models did not substantially outperform the baseline model. It is worth noting that the dimensional models consistently fitted the data better than categorical models with the same K value. However, this could be an artifact of the dimensional models containing more parameters, and more sophisticated measures of model fit are needed to examine this possibility.

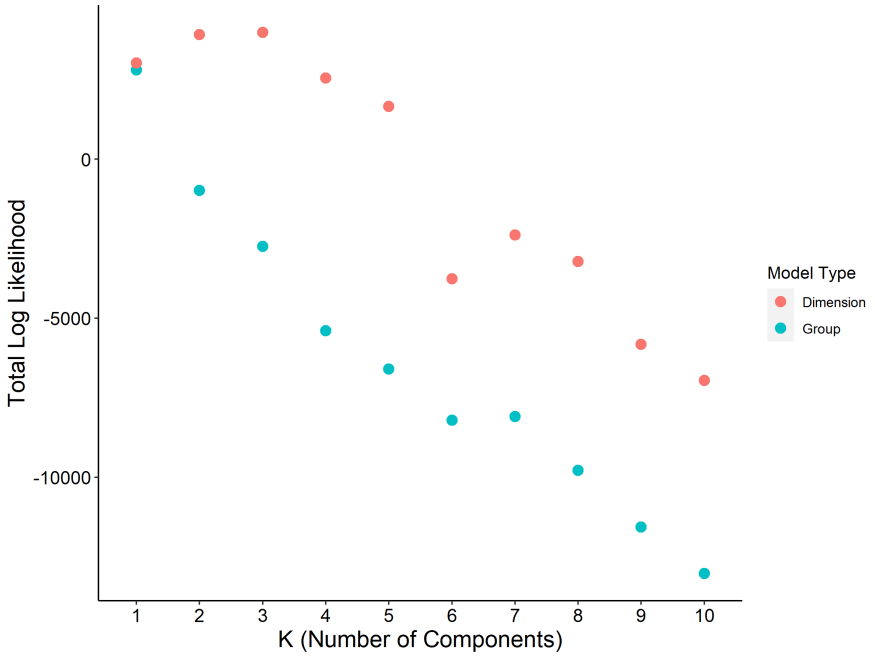

Figure 3.1: Fit of the categorical and the dimensional models as a function of the number of latent components

There are several possible explanations for why latent structure models were not well supported by the data. First, it's possible that neither categorical nor dimensional structures existed in the emotion transition data. However, this explanation is unlikely given

the prevalence of both types of structure across a wide range of domains. A second possibility, however, is that there was too little data for the models to reliably extract the true underlying statistical structure. To examine this possibility, we conducted a model recovery exercise.

Model recovery

In order to further validate that our models indeed captured reliable genuine structures in the data, we conducted a series of model recovery exercises. The exercise consisted of two steps. In the first step, we generated multiple datasets from each of the fitted models. This included both the categorical and the dimensional models, at K values varying from 1 to 10. Datasets were generated at varying sizes ranging from N = 10 to N = 500. In the second step, we re-trained all models, within model type, on each dataset generated by each of the original trained models. As an example to make this procedure more concrete: we first generated multiple datasets of different sizes from the categorical model with $K = 5$. We then trained a series of new categorical models with K values spanning the full range of 1 - 10 on each of these datasets, and computed each model's fit. This process was repeated for each of the 10 original categorical models, as well as each of the dimensional models.

In this exercise, if the models learned non-spurious structures, they should then be recapitulated in the generated data. If the original data had sufficient power for these structures to be learned, generated datasets of similar sizes should be sufficient for the new, refitted models to *recover* the generative K value (identifying it as optimal). If the data did not provide sufficient power, the structures in the models would have been unnecessarily imposed, some of the learned patterns would be spurious. However, these models likely would still

have captured some of the regularities in the original data that would be contained in the generated data. In such a case, using generated datasets with larger sample sizes than the original, we might be able to get a more amplified, more reliable estimate of the true structure.

Categorical Models. The model recovery results for the categorical models are shown in Fig. 3.2. Across all generative K values and the full range of generated sample sizes, the model fits consistently decreased as the K values of the recovery models increased. Such decreases in model fit was particularly pronounced when recovering from larger sized samples. This suggested a high level of confidence that these complex categorical structures was not supported by the data.

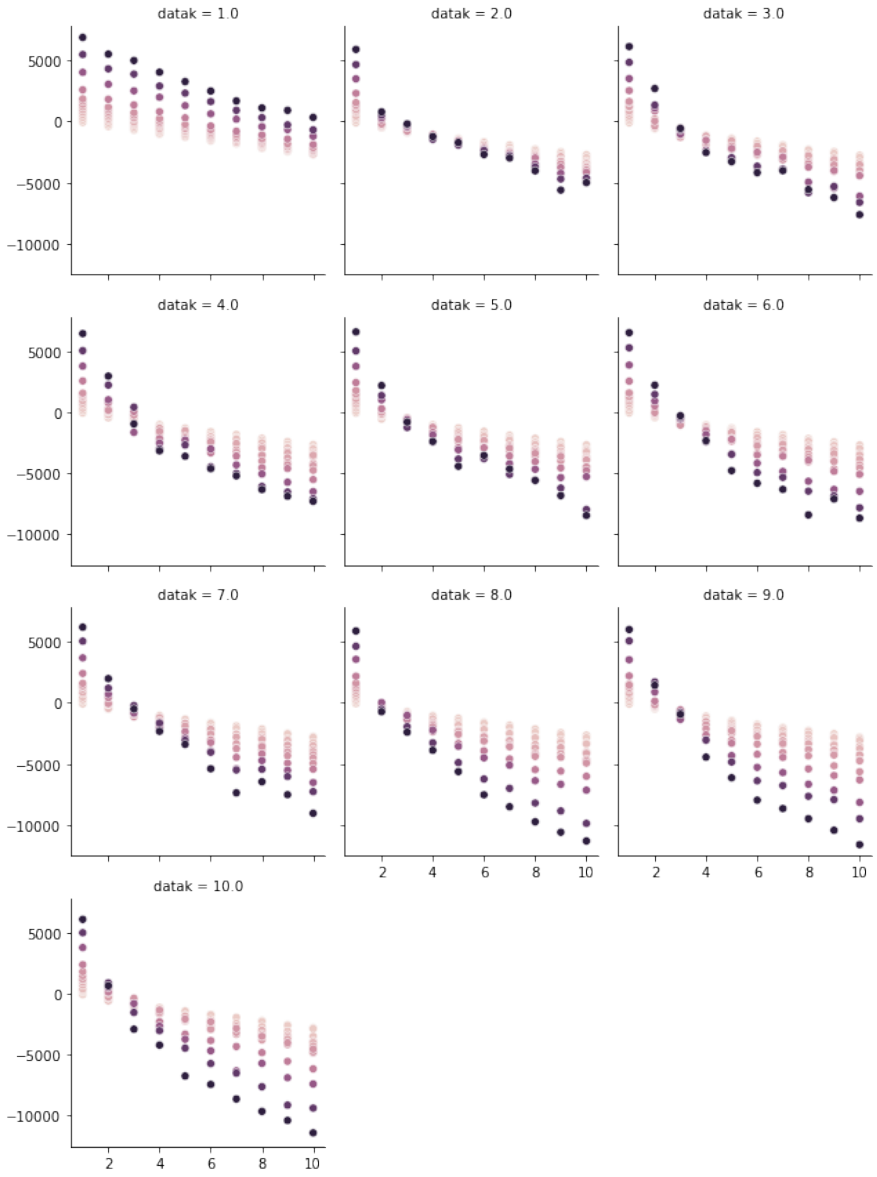

Figure 3.2: Recovering the latent structures learned by the categorical models. Each grid displays the fit of varying-K categorical models on data generated from a single K value. Hues represent the sample size of the data used for refitting

Dimensional Models. The model recovery results for the dimensional models are shown in Fig. 3.3. Two patters are worth noting. First, at larger sample size, the models

consistently fitted the best when the data were generated from a $K = 3$ model. This could be because the $K = 3$ model captured the structure of the original data without learning the spurious patterns learned by models with more latent components. Second, when data are generated from higher dimensional models, recovery with larger sample size consistently identified $K = 3$ as the inflection point. This points to the possibility that the data supported a 3d structure, which was consistently captured by the higher K dimensional models. Both these patterns were consistent with the evidence from the original data suggesting a three dimensional structure underlying emotion transitions. Nonetheless, while the data used for model training was not entirely underpowered, these patterns also pointed out that more data are needed for the modeling results to be definitive.

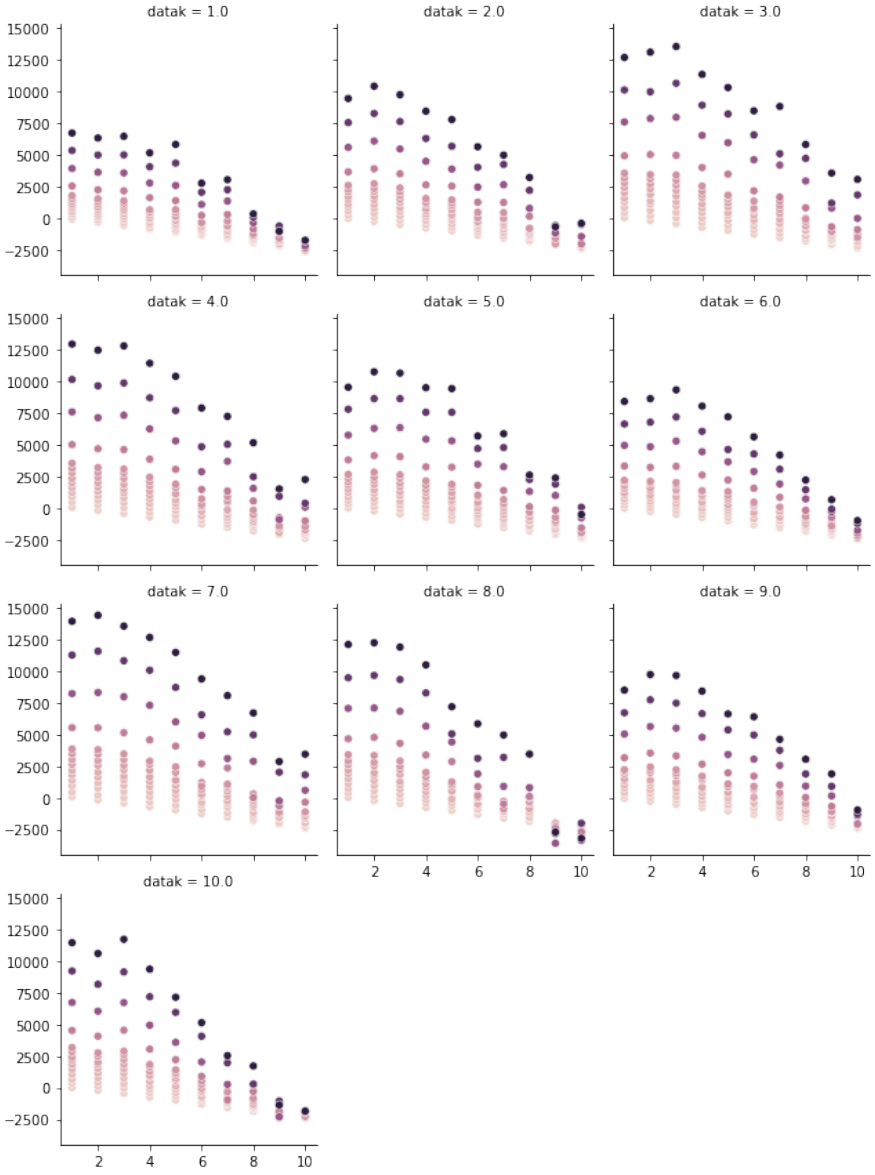

Figure 3.3: Recovering the latent structures learned by the dimensional models. Each grid displays the fit of varying-K dimensional models on data generated from a single K value. Hues represent the sample size of the data used for refitting

3.2 Study 6. Comparing Structured Models Against Baseline

The modeling results from Study 5 suggested, though not definitively, that a dimensional model might better capture the true structure in emotion transitions. To further adjudicate which model better approximated the structured mental representations that people use, we compared model predictions against human judgments on the same task. The following empirical studies capitalized on the fact that, while models with different underlying assumptions can learn from the same data, the structures they learned are different by design. Therefore, they can make divergent predictions given the same input. To determine the validity of the models, we subjected the them to a procedure that simulated an inductive inference task. In this procedure, the models were used to predict emotion transitions based on sparse input. We then identified a set of inputs where the models made the most divergent predictions, and solicited human judgements using those same emotion transition input as stimuli. Finally, we tested model-based hypotheses using human data.

In this study, we aimed to reaffirm, through the modeling approach, that people indeed make inductive inferences about emotion dynamics. The baseline model predicts that people's judgements about unobserved emotion transitions should not change after observing a very small number of transitions. Both the categorical and the dimensional models, on the other hand, suggest that even minimal observations should lead to change in judged probability.

We identify pairs of observed and unobserved transitions for which the latent structure models predicted the most change in predictions as a function of changes in the observed transition. We then used these transition pairs as stimuli in an emotion prediction task,

where participants were asked to predict novel targets' unobserved emotion transitions after receiving sparse observed transitions.

3.2.1 Methods

Participants

Participants (N=102) were recruited from Prolific. All participants were US nationals. One participant was excluded from analyses for failing the comprehension check, leaving a final sample size of 101 participants (mean age = 31.19, 53 female). Participants in this and all subsequent studies provided informed consent in accordance with the Princeton University Institutional Review Board.

Task

Participants completed an emotion prediction task. The task consisted of two types of trials, target learning trials and general prediction trials. On each of the target-learning trials, participants were first introduced to a novel target (e.g. Alex), and were told that the target was based on a participant from past studies. They then read a short sentence describing the target's tendency on a single emotion transition. This *observed* transition was presented as a quote from the target (e.g. "Alex: 'whenever I feel talkative, I am very likely (or unlikely) to feel irritable next"'). They were then asked to rate the likelihood of a separate, *prediction* transition, for that target (e.g. "Now imagine Alex is currently feeling love, how likely will they feel contempt next?"). Furthermore, the target learning trials were divided into two within-subject conditions, a high and a low condition. In the high condition, both the categorical and the dimensional models predicted that participants presented

with the likelihood of the observed transition would rate the inference transition as likely. In the low condition, however, learning the likelihood of the observed transition should lead the participants to rate the inference transition as unlikely.

In addition, we included general prediction trials to measure participants' prior for each emotion pair, without learning learning any information about the specific target. In the general prediction trials, participants were asked to rate the likelihood that people in general would experience certain emotion transitions. The transition items used in the general trials were drawn from the same pool as those in the target learning trials, but the inference transitions were presented for rating without the corresponding stimulus transitions.

A total of ten observation-prediction pairs were used in this study. For each participant, four pairs were randomly selected without replacement. One of the four pairs is used for their target learning trial in the high condition, one for the low condition, and two for the general trials. Each participant rated four trials in total.

Generating Model Predictions

In order to test the models against human judgments, we systematically generated predictions from both the categorical and the dimensional models using sparse sets of emotion transitions as input.

Categorical and dimensional models containing three latent components (K = 3) were trained on one of the independent samples in the dataset (N = 98). In the case of the categorical model, this choice of K was consistent with the nonparametric clustering in Study 5, which discovered three to four meaningful clusters. In the case of the dimensional model, this decision was consistent with past research as well as the results of Study 5, indicating

that three dimensions can successfully capture the representational structures across domains of social cognition (Thornton & Tamir, 2017; Koch et al., 2020). Predictions were made using inputs that only contained a single transition. The set of single transition inputs contained each of the 60 total transitions with all values ranging from 10% to 90% at 10% intervals. To actually generate predictions from the models, we first blocked the models from further updating their learned parameters. The models were then conditioned on an input. The categorical model classified the input as coming from one of the three underlying clusters it had learned. Since each cluster contains its own parameters for all transitions, the model was then able to use those parameters to predict the probability of each unobserved transition, given the small number of input it observed. The dimensional model, on the other hand, computed component mixture weights that likely generated the observed input. It then combined the underlying components by these weights, and used the resulting parameters to predict the unobserved transitions.

In total, the generated predictions contained the expected value of each of the 60 transitions from when the models observed only one transition at a value between 10% and 90%. Stimuli for the behavioral experiments will be selected from these two sets of predictions.

Stimuli Selection

The current study aimed to pitch the two latent structure models against the baseline model, and test the hypothesis that people can indeed learn a novel target's emotion dynamics through a process of inductive reasoning. That is, we must demonstrate that a small amount of observation about a target would alter how people judge the probability of their emotion transition. To maximize the possibility of observing such an effect, we selected

observation-prediction pairs from the set of single-feature predictions such the same amount of change in the stimulus transition led to the largest amount of predicted change in the inference transition. For each potential stimulus transition S, we calculated the predicted difference for all inference transitions between when S is 0.2 and S is 0.8. We then sorted all observation-prediction pairs based on predicted change in descending order, and selected the top ten pairs. We restricted each transition to appear only once as the stimulus and once as the inference, and prohibited a transition to be paired with itself or its inverse (A-B stimuli would not be paired with A-B or B-A inferences).

Analyses

We hypothesized that people learn about others' emotion dynamics through inductive reasoning. That is, differences in initial observed emotion transitions, even when very limited, should lead perceivers to make different predictions about a target's unobserved transitions. In the current study, more specifically, we predicted that participants would rate the emotion transitions in the high condition differently (as more likely) than than those in the low condition. To evaluate this prediction, we conducted a linear mixed effects model regressing transition probability ratings onto condition. Since the high and the low conditions were within subject, we included random intercepts for the participants. Since each observation-prediction pair could be assigned to both conditions, and each pair was rated by multiple participants in each condition, we included both random intercepts and random slopes of condition for the items.

$$RatedProbability = Condition + (1|Participants) + (1 + Condition|Items)$$

While the stimuli were not optimized to distinguish between the categorical and the dimensional models, we additionally explored which one of the two models better approximated human judgment. To do so, we regressed standardized human ratings onto the interaction between model type and standardized model predictions in a linear mixed effects model. We included random intercepts for the participants as well as the items.

$$RatedProbability = ModelPrediction * ModelType + (1|Participants) + (1|Items)$$

3.2.2 Results

Using a linear mixed effects model, we tested whether transition probabilities rated by the participants differed between the high and the low conditions (Fig. 3.4). As predicted by the latent structure models, we found that ratings in the high conditions were significantly higher than ratings in the low condition ($B = 21.72$, $\beta = 0.39$, $t = 4.72$, $p < 0.01$). Variations in a single observed emotion transition led people to make differing predictions about a novel target's unobserved transitions. This suggested that people in fact reason inductively from limited information when they learn another person's emotion dynamics. Furthermore, the effect we detected was in the same direction as predicted by the latent structure models, suggesting that overall the models were able to capture at least some of the latent structures that people employed.

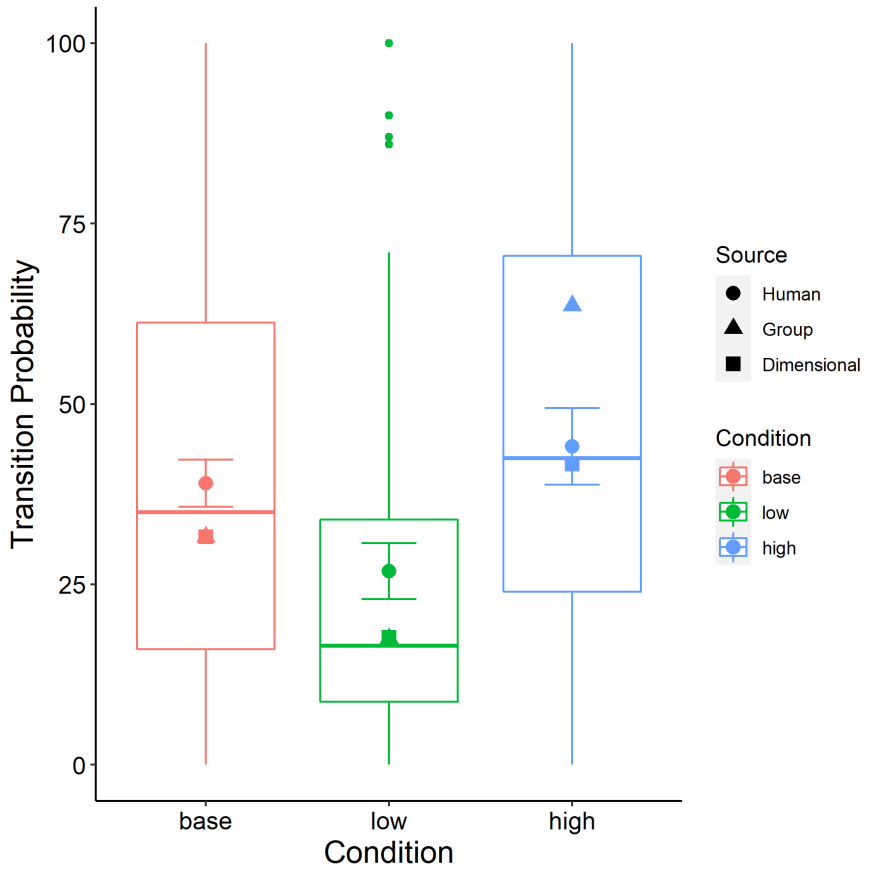

Figure 3.4: Mean transition probability from human judgments and model predictions

Furthermore, we tested whether the categorical model or the dimensional model better approximated people's mental structures (Fig. 3.5). We found a significant interaction between model inferences and model type in predicting human judged transition probability ($B = 0.20$, $\beta = 0.13$, $t = 3.48$, $p < 0.001$). More specifically, inferences made by both models positively predicted human judgments. However, the dimensional model ($\beta = 0.37$) predicted human judgments to a greater extent compared to the categorical model ($\beta = 0.24$). The results suggested that the dimensional model might better capture the mental structure with which people reason inductively about novel social targets'

emotion dynamics.

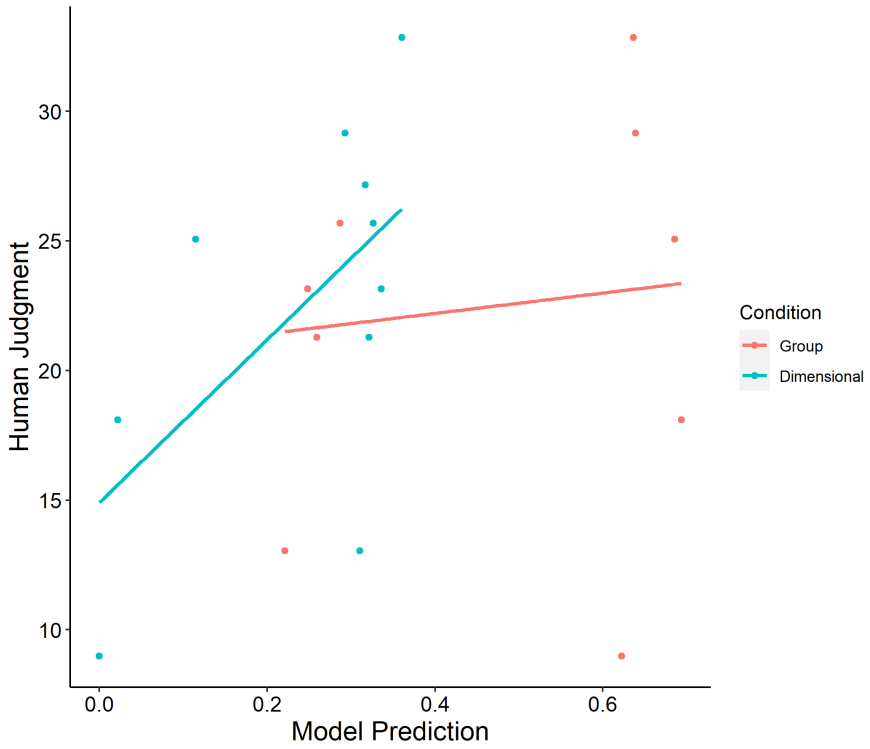

Figure 3.5: Bivariate relations between model-predicted and human-rated differences between the high and the low conditions

3.2.3 Discussion

In a first behavioral study, we established that people in fact reason inductively about the emotion dynamics of novel social partners. Even when they are only given very limited information about a target's emotion transitions, people modify their target-specific judgment about unobserved emotion transitions away from their knowledge about general regularities. Our probabilistic models successfully predicted both the existence and the directions of these adjustments, providing initial evidence that the models captured the latent

structures that guided inductive reasoning about emotion dynamics.

Furthermore, we found preliminary evidence that human judgments tracked more closely with predictions generated from the dimensional model. This finding is consistent with Study 5, where the results suggested that dimensional models with a small number (e.g. 3) of latent components best described the ground truth transition data. This correspondence shed some insight into the consistency between structured mental representations and the true statistical structure of emotion transitions in the population.

3.3 Study 7. Dense Sampling of Transition Space

In Study 6, we presented initial evidence that people reason inductively about emotion dynamics, and that the mental structures that enable this induction is captured by our probabilistic models. Nonetheless, this study had a number of limitations. First, the stimuli used in the study only included a small fraction of the possible observation-prediction pairs. We cannot confidently generalize our results on inductive inference to the vast array of emotion states between which people can transition.

Second, the included stimuli were selected to maximize the predicted differences between the baseline and the latent structure models. The study tested only one of the criteria that the models should be tested against, namely that the model predicts a difference where there actually is a difference. However, all observed emotion transitions cannot be equally informative. While people should adjust their predictions after observing *some* emotion transitions,other transitions should exert little to no influence on emotion predictions. If the model is successful in capturing inductive emotion prediction, it should also be able to predict these negative instances.

To address these limitations, participants completed the emotion transition task from Study 6 with a greatly increased number of stimuli. This allowed us to more densely sample the space. This allowed us to more densely sample the space of possible observation-prediction pairings. We also sampled the stimuli space uniformly with regard to predicted model-baseline differences. That is, pairings where the observation should lead to much or little model-baseline difference were equally likely to be selected as stimuli, allowing us to test the model with both positive and negative instances. Finally, given the results from Studies 5 and 6, we prioritized the difference between baseline and the dimensional model in constructing the current stimuli set. This allows us to more comprehensively test our primary hypothesis that the dimensional model can capture inductive emotion predictions. In addition, we rely on the variation in between-model differences inherent to a large stimuli set to test our secondary hypothesis that the dimensional model should approximate inductive emotion prediction to a greater extent than the categorical model. The hypotheses and analyses for this study were preregistered.

3.3.1 Participants

Participants (N=167) were recruited from Prolific. All participants were US nationals. Four participants were excluded from analyses for failing an attention check, leaving a final sample size of 163 (mean age = 33.45, 93 female). Participants provided informed consent in accordance with the Princeton University Institutional Review Board.

3.3.2 Task

Participants completed an emotion transition learning task. The task is similar to the target learning trials in the emotion transition task from Study 6. The task consisted of nine blocks. In each block, participants were first introduced to a novel target (e.g. Alex) and were told that the target was based on a participant from past studies. They then read a short sentence which described how likely the target was to experience a single emotion transition. This *observed* transition was presented as a quote from the target (e.g. "Alex: 'whenever I feel talkative, I am very likely (or unlikely) to feel irritable next' "). They were then presented with a series of ten transitions (*prediction* items), and asked to rate the likelihood of each for the given target. For each participant, both the observation items and the corresponding prediction items were randomly selected from a larger pool of stimuli, the construction of which we discuss below. Target names were randomly paired with observation items. Each participant provided 90 ratings in total.

3.3.3 Stimuli Selection

In the current study, we aim to more comprehensively test how well the dimensional model approximates the knowledge structures that enable people to reason inductively about emotion transitions. We do so by selecting stimuli that 1) better span the full space of possible observation-prediction pairs, and 2) include stimuli where the observation item was predicted to have a strong influence on the predictions item, as well a stimuli where the observation item was predicted to have little to no influence on the prediction item.

To satisfy these criteria, we first generated single-feature model predictions by applying the same procedure used in Study 6 to the entire dataset. Then, for ease of presentation in

the behavioral task, we first binned all observation-prediction pairs into a high and a low condition based on the probability of the observation item given to the model. The mean of the predictions from when the input was 10%, 20%, and 30% was taken to be the low condition. The mean of the predictions from when the input was 70%, 80% and 90% was taken to be the high condition. With the resulting set of observation-prediction pairs, we split the full range of model predicted difference scores (between the dimensional model and the baseline) into decile bins. We evenly sampled 72 pairs from each bin, and then randomly sampled additional items for input items that were paired with less than 10 prediction items until all inputs are paired with at least 10. We based the selection process on predicted model-baseline differences, as these values represent the predicted *informativeness* of the observation item for the prediction item. The binning process ensured that the stimuli pool evenly spanned the full range of difference scores. The resulting stimuli pool contains 111 unique input items and a total of 1200 observation-prediction pairs. For each participant we randomly selected nine observation items from this pool. We then randomly selected 10 prediction items for each of the nine, resulting in a total of 90 trials.

3.3.4 Analyses

Primarily, we hypothesized that the dimensional model successfully captures the statistical structures enabling people to reason inductively about emotion dynamics from a small number of observations. That is, predictions made by the dimensional model when given a single observation item should track human judgments on the emotion transition learning task. To test this hypothesis, we carried out two linear mixed effects regressions. In the first model, we regressed human ratings onto the dimensional model's predictions. We included

random intercepts, as well as random slopes of model prediction, at both the participant level and the observation item level.

$$RatedProbability = ModelPrediction + (1 + ModelPrediction|Participants) + (1 + ModelPrediction|Items)$$

In the second model, we test the same hypothesis using deviations from the baseline instead of raw scores. We computed human rated differences by subtracting the baseline model's predictions from human ratings on the transition learning task. We computed model predicted differences by subtracting the baseline model's predictions from the dimensional models' predictions. We then regressed human rated differences onto model predicted differences, and included random random intercepts, as well as random slopes of model predicted differences, at both participant and observation item levels.

$$RatedDifference = PredictedDifference + (1 + PredictedDifference|Participants) + (1 + PredictedDifference|Items)$$

We also tested a secondary hypothesis that the dimensional model should more closely track human inductive reasoning compared to the categorical model. We did so by conducting two linear mixed effect models, using raw ratings and model predictions as well as deviations from baseline. In the first model, we regressed human rated transition probabilities onto model predictions, model type, and interaction between the two. At the participant level, we included random intercepts and random slopes of model predictions. At the observation items level, we included random intercepts, random slopes of model predictions, and random slopes of model type.

$$RatedProbability = ModelPrediction * ModelType + (1 + ModelPrediction|Participants) + (1 + ModelPrediction + ModelType|Items)$$

In the second model, we regressed human rated differences from baseline onto model predicted difference from baseline, model, and the interaction between the two. For participants, we included random intercepts and random slopes of model predicted differences. For observation items, included random intercepts, random slopes of model predicted differences, and model type.

3.3.5 Results

Using linear mixed effect models, we tested whether the dimensional model accurately predicted how people make inductive inferences about emotion dynamics from a single observed emotion transition. We found that the predictions made by the dimensional model positively predicted human judgments on the transition learning task ($\beta = 0.38, t(216.72) = 15.52, p < 0.001$; Fig. 3.6). That is, how people predicted a novel target's emotion dynamics after minimal observation conformed to the prescriptions made by a model assuming a dimensional structure. We also found that the predicted differences between the dimensional model and the baseline model positively predicted predicted the difference between human judgments and the baseline ($\beta = 0.28, t(122.63) = 8.29, p < 0.001$; Fig. 3.7). That is, how people adjusted from their general knowledge of emotion dynamics after learning about a specific target also conformed to the prediction of the dimensional model. Together, these results suggested that human inductive learning in the domain of emotion dynamics potentially rely on dimensionally structured underlying knowledge.

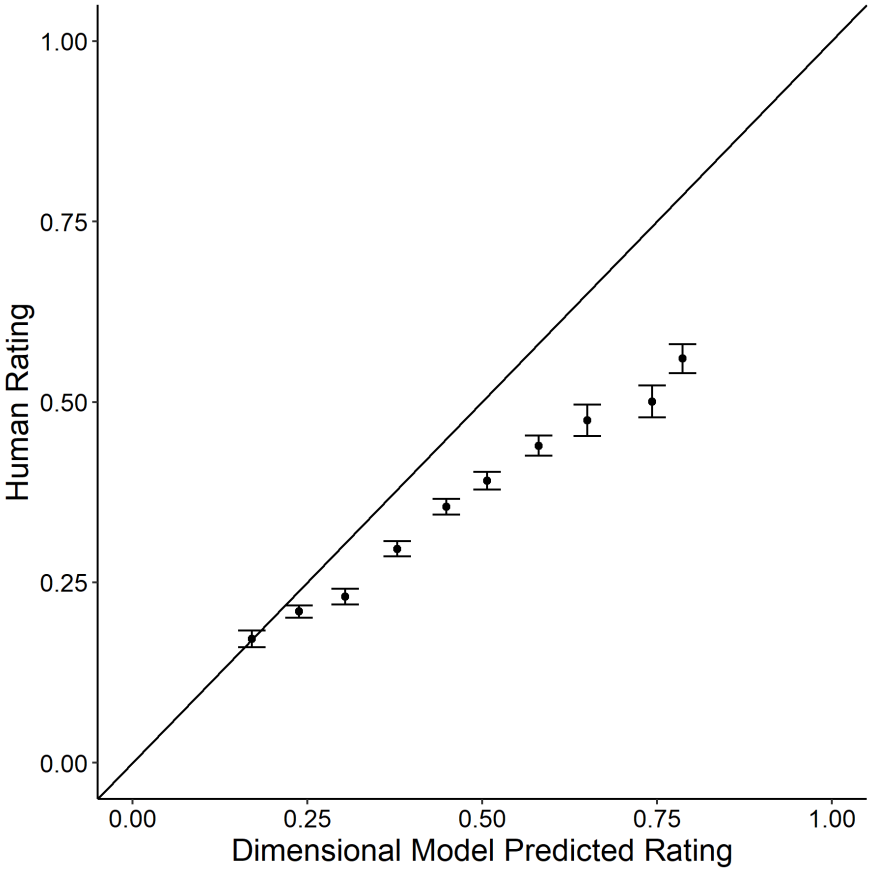

Figure 3.6: Human ratings on the transition learning task tracked the predictions of the dimensional model

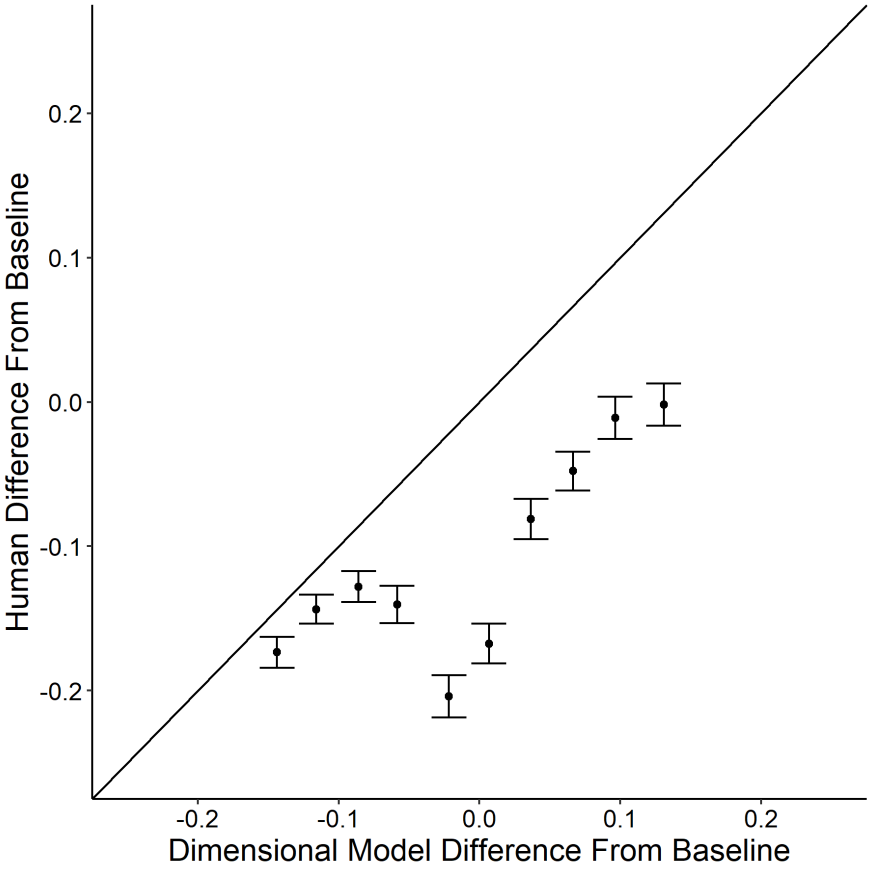

Figure 3.7: Differences between human ratings and the baseline model tracked the differences between the dimensional model and the baseline model

We also tested whether the dimensional model better predicted human judgments than the categorical model. In a linear mixed effects model, we did not find evidence that model type moderated the association between model prediction and human ratings ($\beta = 0.001$, $t(2883.82) = 8.29, p > 0.90$). However, model type did moderate how strongly model-baseline differences predicted human-baseline differences. Differences between the baseline model and the dimensional model, rather than the categorical model, more strongly predicted human-baseline differences ($\beta = 0.09, t(403.89) = 6.01, p < 0.001$). These results

further suggested people might rely on dimension-like, rather than category-based, mental organizations to learn inductively about emotion dynamics.

3.3.6 Discussion

Using a larger quantity of stimuli that more comprehensively spanned the possible space of emotion transitions, we provided further evidence that people reason inductively about others' emotion dynamics. After learning information about a single emotion transition from a novel target, people update their predictions of how likely the target would experience emotion transitions they have not otherwise observed. Furthermore, the dynamic of this learning was captured by a probabilistic model assuming a dimensional latent structure. The dimensional model positively predicted *how* people would update their target-specific beliefs from minimal input. This result suggested the possibility that people's inductive learning in the domain of emotion transitions is guided by underlying mental representations that are organized into a dimensional space.

3.4 Discussion

Emotions are not isolated states. Rather, people's thoughts and feelings unfold over time. Understanding how someone's emotions might change in the future can facilitate social planning help people establish meaningful relationships.

While the people transition between emotion states following general patterns (Thornton & Tamir, 2017), emotion dynamics can also vary widely across individuals and social groups. In order to accurately predict each social partner's emotion changes, people must go beyond simply applying their knowledge about general emotion dynamics, and adjust their

predictions in a partner-specific manner. People interact with many social partners only infrequently, yet the space of possible emotions and emotion transitions is vast. The observations people accrue during limited interactions might not support them in building partner-specific mental models from the ground up. In addition, accurate emotion prediction is time-sensitive, and people need to make partner-level adaptations quickly. If people make inaccurate predictions about their partners, and thus fumble during social interaction, in the early stages of a relationship, it might preclude that relationship from developing further. These two features of naturalistic social interactions make such partner level adaptations particularly challenging.

In the current work, we combine computational models and behavioral experiments to demonstrate how people can address this challenge. Using an emotion transition learning task, we mimic the task of social learning in early-relationship scenarios, where information about a novel partner is scarce. We demonstrate that people can rapidly learn partner-specific emotion dynamics through inductive reasoning, and that they do so by relying on latent statistical structures of how emotion dynamics vary across individuals within the population.

We show that probabilistic models assuming a dimensional structure outperformed those assuming a categorical structure in capturing the variability in emotion transitions across individuals. That is, people cannot be neatly categorized into a small number of groups that are each largely homogenous in terms of emotion dynamics. Rather, people varied in how their emotions change over time along a small number of shared, continuous dimensions.

We also show that people capitalize on such dimensional structures to learn, rapidly and

inductively, how a new social target's emotions might change over time. After observing the likelihood for a previously unknown target to transition between even only a single pair of emotion states, people can adjust their predictions about how likely the target would transition between emotion states that they have not observed. This suggests that people understand how different features of emotion dynamics might be related to each other, and can generalize from scarce observations to unknown properties. Furthermore, we show that the way people adjusted their emotion predictions tracked with the dimensional model's predictions when it is given the same input. This result has two implications. First, it suggests that inductive reasoning about emotion dynamics might be guided by mental representations that are dimensionally organized. This is in line with much past work in social cognition suggesting that low dimensional spaces are a pervasive form of mental representation. Second, as the probabilistic models learned the *actual* statistical structures of emotion transitions, model predictions can be seen as prescriptions of how people *should* rate the probabilities in the transition learning task. The congruence between human judgments and model predictions thus suggests that people in fact encode these structures and apply them for inferences, and that inductive learning in the domain of emotion dynamics is, to a certain extent, normatively accurate.

While we demonstrate that informative prior knowledge can help people quickly become accurate about novel targets' emotion dynamics, our research leaves open the question of how these structured representations are acquired in the first place. One possibility is that people acquire these structures through cumulative statistical learning over the course of social development. From a young age, children can learn the statistical dependencies in a sequence of events through bottom-up observations. This ability is most

prominently demonstrated in cognitive domains such as language acquisition (Newport & Aslin, 2004b) and visual perception (Fiser & Aslin, 2002). Importantly, statistical learning mechanisms generalize to more social phenomena. In addition to emotion transitions, people can also learn and utilize the transitional probabilities between actions in a sequence (Monroy, Gerson, & Hunnius, 2017; Thornton & Tamir, 2021). Distributional properties shape how people associate facial features with personality traits (Dotsch, Hassin, & Todorov, 2016), and it has been proposed that basic statistical learning might underlie the emergence of complex social cognition such as theory of mind (Ruffman, Taumoepeau, & Perkins, 2012). It is possible that over a large quantity of social observations, statistical learning can also support people to extract complex, hierarchical structures of emotion transitions from scratch.

However, it has been argued that the learning of complex statistical structures cannot be supported by bottom-up, associative processes alone (Gold, 1967). In order to successfully discover the hierarchical structures of how emotion dynamics might vary across individuals, the learning process need to be scaffolded by more abstract, more domain general knowledge about the world, such as inductive biases over possible statistical structures (Gershman & Niv, 2010). People often face a wide array of inductive learning problems. Many of these problems might share underlying solutions, and statistical structures acquired in one domain can often be transferred to other domains to facilitate learning (Kemp, Goodman, & Tenenbaum, 2010). It is likely that bottom-up and top-down learning processes both contribute to people's structured knowledge about emotion dynamics, and future work should further explore the emergence of such knowledge throughout the time-course of development.

At a much more local level, bottom-up statistical learning and top-down inductive learning might also jointly enable people to adapt their emotion predictions to a specific partner. The current work formalizes only the very early stage of relationship development, where partner specific information is scarce. Under such a scenario, people might need to lean more heavily on prior knowledge and inductive processes to rapidly refine their inferences about the partner. However, in more long-standing interpersonal relationships, people often accrue large quantities of experiences with their partner that it is possible to form a partner-specific mental model in a more bottom-up manner. In keeping with this hypothesis, rich partner-specific knowledge is often strongly associated with predictive accuracy in close friends (Stinson & Ickes, 1992; Zhao et al., 2020). Future research should aim to characterize the dynamics of partner-specific learning and belief updating over the course of relationship development.

The current work adds to the growing literature using probabilistic models to study cognition in the social domain. The probabilistic modeling approach enables researchers to flexibly formalize hypotheses across a wide range of social phenomena. To name a few, it has been used to quantify how people integrate private information with social advice (Vélez & Gweon, 2019), how people understand others' minds through intuitive theories of rational utility maximization (Jara-Ettinger, Schulz, & Tenenbaum, 2020), and how people can form norms and conventions across multiple social scales (Hawkins, Goodman, & Goldstone, 2019; Hawkins et al., 2021). Much previous research have constructed well-modeled laboratory experiments using stripped down, well controlled artificial stimuli, and can therefore often make precise predictions that correlate highly with human behaviors ($r = 0.95$; Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017). In comparison, the current

work is an effort to scale up such methods. We studied spontaneously occurring dyads, and intentionally sampled large quantities of emotion transitions to cover the stimuli space as fully as possible. The decision to study people's understanding of emotion dynamics in its naturalistic context has several implications for modeling. First, there is a greater probability of model mis-specification. While much past research has focused on the dichotomy between categorical and dimensional representations of mental state concepts (e.g. Cowen & Keltner, 2017; Tamir et al., 2016), it is possible that neither provides the best fit for how people represent emotion dynamics. Recent work has demonstrated that probabilistic modeling can be used to adjudicate between representational forms (Kemp & Tenenbaum, 2008). Future work modeling emotion dynamics can benefit from taking a similar approach and selecting candidate latent structures in a more principled, data driven manner. Second, our model focuses on only the latent statistical structure of emotion dynamics, and inevitably leaves out other sources of variance. For example, we assume that people hold the same general knowledge about emotion dynamics and rely on the same latent structure, whereas they can vary widely on both. A direct consequence is that our model predictions did not achieve the same level of correspondence with human judgments ($\beta = 0.38$) as did previous work. Future work can integrate these additional variances by increasing within-subject response sampling and establishing participant-level baselines in order to boost the precision of the model.

Finally, it is worth pointing out that the current modeling work takes a data-driven approach that focuses on *prediciting* human behavior on a given task. While the hypotheses about possible structures are theoretically derived from past work, the contents that populated these structures (e.g. the category and dimension components) were discovered com-

putationally. In order to achieve deeper psychological insights, it is imperative that future modeling work attempt to interpret computational findings in tandem with cumulative verbal theory, and that future theoretical work in social and affective sciences take into account analyses at the algorithmic level.

4

Conclusion

In this dissertation, my primary goal has been to situate the phenomenon of emotion prediction in its naturalistic social and relationship contexts. Across three projects, I investigated how people accurately understand emotion dynamics and predict emotion transitions at a partner-specific level, given the information constraints imposed by different stages of relationship development.

Chapter 1 aimed to establish that emotion prediction is an ecologically valid and relevant facet of day-to-day social cognition. Results demonstrated that people can accurately pre-

dict emotion transitions at a partner-specific level. Their levels of accuracy was related to how familiar they were with their target, as well as the amount of person knowledge they possessed. Furthermore, partner-specific emotion prediction showed both divergent vailidity, diverging from previously studied domains of social cognition such as emotion recognition, and predictive validities, predicting downstream implications for people's success in their social lives.

Chapter 2 investigated the mechanisms of emotion prediction when people are faced with informational constraints imposed by relationship contexts. In the absence of partner-specific information, self-knowledge can be a highly effective proxy for accurate emotion prediction, likely due to the high degrees of homophily in naturalistic social networks. In addition, how much people leaned on self-knowledge to predict others' emotion transitions might be tuned to their chronic social positions. More typical individuals learn to make less partner-specific adaptations at the outset, and as a result more accurately predict others' emotions.

Finally, Chapter 3 investigated the learning process through which people rapidly individuate their emotion predictions to novel partners. People not only understand the general dynamics of emotions, but they also understand how emotion dynamics might vary across individuals in the population. As soon as people accumulate a small amount of initial information about a novel target's emotion dynamics, they can exploit such statistical structures to quickly infer their partner's unobserved emotion dynamics.

Findings from these projects contribute to the the social psychology literature by expanding our understanding of the human capacity for "mind-reading", suggesting that not only can people make complex inferences about other minds in the moment, but they also

understand how others' experiences ebb and flow over time, and can use this understanding to predictively see into others' emotional future. Mentalistic concepts such as goals, beliefs, desires, and emotions are a useful tool. People rely on these notions of the mind to understand others' behaviors. Due to its critical importance for social functioning, research from both social and cognitive psychology have focused on how people understand others' mental states. However, theory of mind research has largely focused on the questions of in-the-moment inferences, where people use temporally adjacent, often observable information such as facial expressions (Wu, Baker, Tenenbaum, & Schulz, 2018), tones of voice (Wu, Muentener, & Schulz, 2017), and features of surrounding events (Skerry & Saxe, 2015) to derive others' mental states. The present work suggests that people's mind-reading ability extends to information that is more temporally distant and less rooted in momentary perception. People understand how people's emotions change over time, as well how the dynamics of these changes can vary from partner to partner.

Knowledge about emotion dynamics enriches momentary mental state inferences in at least two ways. On one hand, it means that information about other's previous states can serve as an additional source of information that helps people contextualize their observations and figure out how others might feel. For example, people might have very different interpretations of the sound of someone crying if they were told the target was recently feeling happy, as opposed to distraught.

More importantly, it boosts the utility of mental state understanding as the basis for *predictions* at a much longer time scale. Conventional theorizing identifies notions of the mind as a powerful tool for short range predictions. Angry people aggress, whereas amused people relax. People intuitively grasp these immediate consequences of mental states. However,

when long-range emotion dynamics are taken into account, knowing how someone feels at a given moment gives people information about how they might feel in the future. This allows people to plan their actions strategically farther ahead in time.

It is worth noting, however, that the dichotomy between long-ranging emotion dynamics and momentary emotion recognition is artificial, and the two phenomena likely interact in complex ways to form the full picture of mental state inference. In the current work, emotion dynamics are construed as a Markov process. This entails a simplifying assumption –that a person's next emotion state depends only on their current state. However, transitions between different emotions can be mediated through multiple pathways. These pathways can be internal to the person, such as through rumination, mind wandering, and emotion regulation. They could also be external –emotions can lead people to seek out specific types of situations and interactions (Zaki & Williams, 2013), which in turn lead to different subsequent emotion states. It is possible that people understand emotion dynamics because they have mental models of these intermediate causal events. One major avenue for future research is to relax current Markovian assumptions and build out the complexity of emotion prediction by exploring people's understanding of how emotion transitions take place. For example, researchers can investigate folk theories of emotion regulation –do people have systematic expectations of how others deal with different kinds of emotions? Do the same intermediate steps lead to different resulting emotions when the preexisting states differ? Ultimately, these the goal should be to understand how different "layers" of the social world interact dynamically, both in ground truth and in mental models (Tamir & Thornton, 2018).

The current dissertation also contributes to the literature by demonstrating that social

cognition can, and should, be studied in its natural environment. In daily social interactions, people face constraints that can be easily overlooked by simplified, controlled laboratory experiments. In our case, emotion prediction happens at a partner-specific level, embedded in a social network, and at a specific point in relationship and developmental time. Each of these contextual factors affects the information available to the perceiver, as well as their optimal inferential strategy. That is, the umbrella phenomenon we refer to as emotion prediction is, in fact, several different computational problems "in a trench coat." Pluralism at the computational level begets pluralism at the algorithmic level. As I have shown across these projects, people can make emotion predictions using several different strategies, including egocentrism (Chapter 2), inductive learning (Chapter 3), and partner-specific mental models (Chapter 1). And even more importantly, people rationally turn to these strategies to different extents based on social context, often leading their predictions to be accurate.

One major avenue for research into naturalistic social cognition is to look at the dynamical, possibly reciprocal relationship between social cognition and naturalistic contexts. In the domain of interpersonal relationship, researchers can investigate how people might shift the weights they assign to different kinds of learning and different inferential strategies as relationships develop (as discussed in Chapter 3). At a more macro scale, researchers investigate how social cognition and social networks shape each other. For example, it is possible homophily, a well-known characteristic of human social networks, arises from constraints on individual level social cognition. Namely, there might be a ceiling to how accurately people can make social predictions in the absence of a high level of interpersonal similarity, such that relationships between dissimilar individual are terminated at an early

stage due to unsuccessful, inaccurate mind reading efforts. Future studies can use methods such as agent-based modeling to formalize how cluster structures in social networks might be a result of individuals jointly maximizing the accuracy and efficiency of their predictions.

Overall, the works presented in this dissertation demonstrated the importance of considering any cognitive and psychological phenomenon not as an unitary entity, but as a plurality of computational problems. The complexity of the mind did not arise in response to simple challenges, and its geniuses could only be fully understood when it's analyzed alongside the complexity of the problems it faces. This dissertation amounts to an attempt at studying the social mind as it is situated in the society of minds, revealing the mechanism and utility of predictive social cognition. Future research should aim to extend this holistic approach to paint a fuller and fuller picture of how we manage to navigate this whirling, buzzing social world.

References

Alicke, M., Mandel, D., Hilton, D., Gerstenberg, T., & Lagnado, D. (2015). Causal conceptions in social explanation and moral evaluation: A historical tour. *Perspectives on Psychological Science*, *10*(6), 790–812. Retrieved from https://doi.org/10.1177/1745691615601888 doi: 10.1177/1745691615601888

Ambady, N. (2010). The perils of pondering: Intuition and thin slice judgments. *Psychological Inquiry*, *21*(4), 271–278. Retrieved from https://doi.org/10.1080/1047840x.2010.524882 doi: 10.1080/1047840X.2010.524882

Ames, D. (2004a). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, *87*(5), 573–585. Retrieved from https://doi.org/10.1037/0022-3514.87.5.573 doi: 10.1037/0022-3514.87.5.573

Ames, D. (2004b). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, *87*(5), 573–585. Retrieved from https://doi.org/10.1037/0022-3514.87.5.573 doi: 10.1037/0022-3514.87.5.573

Bahns, A., Crandall, C., Gillath, O., & Preacher, K. (2017). Similarity in relationships as niche construction: Choice, stability, and influence within dyads in a free choice environment. *Journal of Personality and Social Psychology*, *112*(2), 329–355. Retrieved from https://doi.org/10.1037/pspp0000088 doi: 10.1037/pspp0000088

Bahns, A., Springer, L., & The, C. (2015). *Fostering diverse friendships: The role of beliefs about the value of diversity*. Group Processes & Intergroup Relations. Retrieved from http://journals.sagepub.com/doi/10.1177/1368430214566893

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*, 0064.

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development*, *82*(6), 1887–1905. Retrieved from https://doi.org/10.1111/j.1467-8624.2011.01669.x doi: 10.1111/j.1467-8624.2011.01669.x

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. Retrieved from https://doi.org/10.1111/j.1469-7610.1997.tb01599.x doi: 10.1111/j.1469-7610.1997.tb01599.x

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. Retrieved from https://doi.org/10.1093/scan/nsw154 doi: 10.1093/scan/nsw154

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290. Retrieved from https://doi.org/10.1177/0963721411422522 doi: 10.1177/0963721411422522

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bauman, K., & Geher, G. (2002). We think you agree: The detrimental impact of the false consensus effect on behavior. *Current Psychology*, *21*(4), 293–318. Retrieved from https://doi.org/10.1007/s12144-002-1020-0 doi: 10.1007/s12144-002-1020-0

Baumeister, R. F., & Leary, M. R. (2017). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Interpersonal development*, 57–89.

Betz, N., Hoemann, K., & Barrett, L. F. (2019). Words are a context for mental inference. *Emotion*, *19*(8), 1463–1477. Retrieved from https://doi.org/10.1037/emo0000510 doi: 10.1037/emo0000510

Biesanz, J. C. (2010). The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, *45*(5), 853–885. Retrieved from https://doi.org/10.1080/00273171.2010.519262 doi: 10.1080/00273171.2010.519262

Biesanz, J. C., West, S., & Millevoi, A. (2007). What do you learn about someone over time? the relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, *92*(1),

119–135. Retrieved from https://doi.org/10.1037/0022-3514.92.1.119 doi: 10.1037/0022-3514.92.1.119

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., ... Goodman, N. D. (2018). Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*.

Birch, S., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382–386. Retrieved from https://doi.org/10.1111/j.1467-9280.2007.01909.x doi: 10.1111/j.1467-9280.2007.01909.x

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993–1022.

Bosacki, S., & Astington, J. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social Development*, *8*(2), 237–255. Retrieved from https://doi.org/10.1111/1467-9507.00093 doi: 10.1111/1467-9507.00093

Buchsbaum, D., Griffiths, T., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology*, *76*, 30–77. Retrieved from https://doi.org/10.1016/j.cogpsych.2014.10.001 doi: 10.1016/j.cogpsych.2014.10.001

Buote, V., Pancer, S., Pratt, M., Adams, G., Birnie-Lefcovitch, S., Polivy, J., & Wintre, M. (2007). The importance of friends: Friendship and adjustment among 1st-year university students. *Journal of Adolescent Research*, *22*(6), 665–689. Retrieved from https://doi.org/10.1177/0743558407306344 doi: 10.1177/0743558407306344

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*(5), 1232–1254. Retrieved from https://doi.org/10.1086/261651 doi: 10.1086/261651

Campbell, K., Marsden, P., & Hurlbert, J. (1986). Social resources and socioeconomic status. *Social Networks*, *8*(1), 97–117. Retrieved from https://doi.org/10.1016/S0378-8733(86)80017-X doi: 10.1016/S0378-8733(86)80017-X

Cantor, N., & Mischel, W. (1979). Prototypicality and personality: Effects on free recall and personality impressions. *Journal of Research in Personality*, *13*(2), 187–205. Retrieved from https://doi.org/10.1016/0092-6566(79)90030-8 doi: 10.1016/0092-6566(79)90030-8

Cassels, T., & Birch, S. (2014). Comparisons of an open-ended vs. forced-choice 'mind reading' task: Implications for measuring perspective-taking and emotion recognition. *PLOS ONE*, *9*(12), 93653. Retrieved from https://doi.org/10.1371/journal.pone.0093653 doi: 10.1371/journal.pone.0093653

Cho, J., & Knowles, E. (2013). I'm like you and you're like me: Social projection and self-stereotyping both help explain self–other correspondence. *Journal of Personality and Social Psychology*, *104*(3), 444–456. Retrieved from https://doi.org/10.1037/a0031017 doi: 10.1037/a0031017

Clement, R., & Krueger, J. (2002). Social categorization moderates social projection. *Journal of Experimental Social Psychology*, *38*(3), 219–231. Retrieved from https://doi.org/10.1006/jesp.2001.1503 doi: 10.1006/jesp.2001.1503

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38), E7900–E7909.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the bias map. *Advances in experimental social psychology*, *40*, 61–149.

Cunningham, W., Dunfield, K., & Stillman, P. (2013). Emotional states from affective dynamics. *Emotion Review*, *5*(4), 344–355. Retrieved from https://doi.org/10.1177/1754073913489749 doi: 10.1177/1754073913489749

Davidson, R. (2004). Well-being and affective style: Neural substrates and biobehavioural correlates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1449), 1395–1411. Retrieved from https://doi.org/10.1098/rstb.2004.1510 doi: 10.1098/rstb.2004.1510

Dawes, R. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, *25*(1), 1–17. Retrieved from https://doi.org/10.1016/0022-1031(89)90036-X doi: 10.1016/0022-1031(89)90036-X

Decety, J., & Jackson, P. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. Retrieved from https://doi.org/10.1177/1534582304267187 doi: 10.1177/1534582304267187

Dennett, D. C. (1989). *The intentional stance*. MIT press.

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour*, *1*(1), 1–6.

Eberhardt, J., Davies, P., Purdie-Vaughns, V., & Johnson, S. (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386. Retrieved from https://doi.org/10.1111/j.1467-9280.2006.01716.x doi: 10.1111/j.1467-9280.2006.01716.x

Epley, N. (2004). A tale of tuned decks? anchoring as accessibility and anchoring as adjustment. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (p. 240–257). Blackwell Publishing Ltd. Retrieved from https://doi.org/10.1002/9780470752937.ch12 doi: 10.1002/9780470752937.ch12

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. Retrieved from https://doi.org/10.1037/0022-3514.87.3.327 doi: 10.1037/0022-3514.87.3.327

Essien, I., Otten, S., & Degner, J. (2020). Group evaluations as self-group distancing: Ingroup typicality moderates evaluative intergroup bias in stigmatized groups. *European Journal of Social Psychology*. Retrieved from https://doi.org/10.1002/ejsp.2708 doi: 10.1002/ejsp.2708

Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of Personality and Social Psychology*, *114*(4), 547–571. Retrieved from https://doi.org/10.1037/pspa0000115 doi: 10.1037/pspa0000115

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

Funder, D. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652–670. Retrieved from https://doi.org/10.1037/0033-295X.102.4.652 doi: 10.1037/0033-295X.102.4.652

Furr, R. (2008). *A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness—furr—2008—journal of personality—wiley online library*. Retrieved from https://onlinelibrary-wiley-com.ezproxy.princeton.edu/doi/full/10.1111/j.1467-6494.2008.00521.x

Gendron, M., Roberson, D., Vyver, J., & Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, *14*(2), 251–262. Retrieved from https://doi.org/10.1037/a0036052 doi: 10.1037/a0036052

Gershman, S. J., Gerstenberg, T., Baker, C., & Cushman, F. (2016). Plans, habits, and theory of mind. *PLOS ONE*, *11*(9), 0162246. Retrieved from https://doi.org/10.1371/journal.pone.0162246 doi: 10.1371/journal.pone.0162246

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Current opinion in neurobiology*, *20*(2), 251–256.

Gershman, S. J., Pouncy, H. T., & Gweon, H. (2017). Learning the structure of social influence. *Cognitive Science*, *41*.

Gilbert, D., Pinel, E., Wilson, T., Blumberg, S., & Wheatley, T. (1998). *Immune neglect: A source of durability bias in affective forecasting*. Undefined.

Gilovich, T., Medvec, V., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology*, *78*(2), 211–222. Retrieved from https://doi.org/10.1037/0022-3514.78.2.211 doi: 10.1037/0022-3514.78.2.211

Gilovich, T., Savitsky, K., & Medvec, V. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology*, *75*(2), 332–346. Retrieved from https://doi.org/10.1037/0022-3514.75.2.332 doi: 10.1037/0022-3514.75.2.332

Gleason, K., Jensen-Campbell, L., & Ickes, W. (2009). The role of empathic accuracy in adolescents' peer relations and adjustment. *Personality and Social Psychology Bulletin*, *35*(8), 997–1011. Retrieved from https://doi.org/10.1177/0146167209336605 doi: 10.1177/0146167209336605

Gold, E. M. (1967). Language identification in the limit. *Information and control*, *10*(5), 447–474.

Goldman, L., & Hogg, M. A. (2016). Going to extremes for one's group: The role of prototypicality and group acceptance. *Journal of Applied Social Psychology*, *46*(9), 544–553. Retrieved from https://doi.org/10.1111/jasp.12382 doi: 10.1111/jasp.12382

Gopnik, A., & Wellman, H. (2012). Reconstructing constructivism: Causal models, bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. Retrieved from https://doi.org/10.1037/a0028044 doi: 10.1037/a0028044

Gordon, R. (1986). Folk psychology as simulation. *Mind & Language*, *1*(2), 158–171.

Gow, A., Pattie, A., Whiteman, M., Whalley, L., & Deary, I. (2007). Social support and successful aging. *Journal of Individual Differences*, *28*(3), 103–115. Retrieved from https://doi.org/10.1027/1614-0001.28.3.103 doi: 10.1027/1614-0001.28.3.103

Granberg, D., Jefferson, N., Brent, E., & King, M. (1981). Membership group, reference group, and the attribution of attitudes to groups. *Journal of Personality and Social Psychology*, *40*(5), 833–842. Retrieved from https://doi.org/10.1037/0022-3514.40.5.833 doi: 10.1037/0022-3514.40.5.833

Griffiths, T. (2020). Understanding human intelligence through human limitations. *Trends in Cognitive Sciences*, *24*(11), 873–883. Retrieved from https://doi.org/10.1016/j.tics.2020.09.001 doi: 10.1016/j.tics.2020.09.001

Griffiths, T., Lieder, F., & Goodman, N. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. Retrieved from https://doi.org/10.1111/tops.12142 doi: 10.1111/tops.12142

Gross, J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281–291. Retrieved from https://doi.org/10.1017/S0048577201393198 doi: 10.1017/S0048577201393198

Hawkins, R. D., Franke, M., Frank, M. C., Smith, K., Griffiths, T. L., & Goodman, N. D. (2021). *From partners to populations: A hierarchical bayesian account of coordination and convention.*

Hawkins, R. D., Goodman, N. D., & Goldstone, R. L. (2019). The emergence of social norms and conventions. *Trends in cognitive sciences*, *23*(2), 158–169.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology*, *57*(2), 243–259.

Heller, A., Fox, A., Wing, E., McQuisition, K., Vack, N., & Davidson, R. (2015). The neurodynamics of affect in the laboratory predicts persistence of real-world emotional

responses. *Journal of Neuroscience*, *35*(29), 10503–10509. Retrieved from https://doi.org/10.1523/JNEUROSCI.0569-15.2015 doi: 10.1523/JNEUROSCI.0569-15.2015

Hoch, S. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology*, *53*(2), 221–234. Retrieved from https://doi.org/10.1037/0022-3514.53.2.221 doi: 10.1037/0022-3514.53.2.221

Hogg, M. A. (2020). *Social identity theory*. Stanford University Press.

Holtz, R. (2003). Intragroup or intergroup attitude projection can increase opinion certainty: Is there classism at college? *Journal of Applied Social Psychology*, *33*(9), 1922–1944. Retrieved from https://doi.org/10.1111/j.1559-1816.2003.tb02087.x doi: 10.1111/j.1559-1816.2003.tb02087.x

Human, L. J., & Biesanz, J. C. (2011). Through the looking glass clearly: accuracy and assumed similarity in well-adjusted individuals' first impressions. *Journal of Personality and Social Psychology*, *100*(2), 349.

Human, L. J., Carlson, E. N., Geukes, K., Nestler, S., & Back, M. D. (2020). Do accurate personality impressions benefit early relationship development? the bidirectional associations between accuracy and liking. *Journal of Personality and Social Psychology*, *118*(1), 199.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, *61*(4), 587–610. Retrieved from https://doi.org/10.1111/j.1467-6494.1993.tb00783.x doi: 10.1111/j.1467-6494.1993.tb00783.x

Ickes, W., Robertson, E., Tooke, W., & Teng, G. (1986). Naturalistic social cognition: Methodology, assessment, and validation. *Journal of Personality and Social Psychology*, *51*(1), 66.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*(4), 730–742. Retrieved from https://doi.org/10.1037/0022-3514.59.4.730 doi: 10.1037/0022-3514.59.4.730

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The naive utility calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*.

Judd, C., & Johnson, J. (1981). Attitudes, polarization, and diagnosticity: Exploring the effect of affect. *Journal of Personality and Social Psychology*, *41*(1), 26–36. Retrieved from https://doi.org/10.1037/0022-3514.41.1.26 doi: 10.1037/0022-3514.41.1.26

Jussim, L., Crawford, J., & Rubinstein, R. (2015). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science*, *24*(6), 490–497. Retrieved from https://doi.org/10.1177/0963721415605257 doi: 10.1177/0963721415605257

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*(7), 1185–1243.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687–10692.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, *26*(2), 165–208. Retrieved from https://doi.org/10.1006/cogp.1994.1006 doi: 10.1006/cogp.1994.1006

Keysar, B., & Henly, A. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, *13*(3), 207–212. Retrieved from https://doi.org/10.1111/1467-9280.00439 doi: 10.1111/1467-9280.00439

Klein, S. B., Loftus, J., & Burton, H. A. (1989). Two self-reference effects: The importance of distinguishing between self-descriptiveness judgments and autobiographical retrieval in self-referent encoding. *Journal of personality and social psychology*, *56*(6), 853.

Koch, A., Imhoff, R., Unkelbach, C., Nicolas, G., Fiske, S., Terache, J., ... Yzerbyt, V. (2020). Groups' warmth is a personal matter: Understanding consensus on stereotype dimensions reconciles adversarial models of social evaluation. *Journal of Experimental Social Psychology*, *89*, 103995.

Kolb, K. (2014). Emotional subcultures. *Sociology Compass*, *8*(11), 1229–1241. Retrieved from https://doi.org/10.1111/soc4.12207 doi: 10.1111/soc4.12207

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*(6), 1042–1060. Retrieved from https://doi.org/10.1037/a0020962 doi: 10.1037/a0020962

Li, Y.-Y., & Hong, Q. (2016). Intergroup perceptual accuracy predicts real-life intergroup interactions. *Group Processes & Intergroup Relations*. Retrieved from http://journals.sagepub.com/doi/10.1177/1368430201004004004

Lieder, F., & Griffiths, T. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*. Retrieved from https://doi.org/10.1017/S0140525X1900061X doi: 10.1017/S0140525X1900061X

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

Martinez, L., Falvello, V., Aviezer, H., & Todorov, A. (2016). Contributions of facial expressions and body language to the rapid perception of dynamic emotions. *Cognition and Emotion*, *30*(5), 939–952. Retrieved from https://doi.org/10.1080/02699931.2015.1035229 doi: 10.1080/02699931.2015.1035229

McConnell, A., Sherman, S., & Hamilton, D. (1994). Illusory correlation in the perception of groups: An extension of the distinctiveness-based account. *Journal of Personality and Social Psychology*, *67*(3), 414–429. Retrieved from https://doi.org/10.1037/0022-3514.67.3.414 doi: 10.1037/0022-3514.67.3.414

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.

Mendelson, M., & Aboud, F. (1999). Measuring friendship quality in late adolescents and young adults: Mcgill friendship questionnaires. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *31*(2), 130–132. Retrieved from https://doi.org/10.1037/h0087080 doi: 10.1037/h0087080

Mendelson, M., & Kay, A. (2003). Positive feelings in friendship: Does imbalance in the relationship matter? *Journal of Social and Personal Relationships*, *20*(1), 101–116. Retrieved from https://doi.org/10.1177/02654075030201005 doi: 10.1177/02654075030201005

Mesquita, B., Boiger, M., & De Leersnyder, J. (2016). The cultural construction of emotions. *Current Opinion in Psychology*, *8*, 31–36. Retrieved from https://doi.org/10.1016/j.copsyc.2015.09.015 doi: 10.1016/j.copsyc.2015.09.015

Mildner, J. N., & Tamir, D. I. (2018, Dec). *The people around you are inside your head: Social context shapes spontaneous thought.* PsyArXiv. Retrieved from psyarxiv.com/xmzh7 doi: 10.31234/osf.io/xmzh7

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of experimental child psychology*, *157*, 14–28.

Morelli, S., Ong, D., Makati, R., Jackson, M., & Zaki, J. (2017). Empathy and well-being correlate with centrality in different social networks. *Proceedings of the National Academy of Sciences*, *114*(37), 9843–9847. Retrieved from https://doi.org/10.1073/pnas.1702155114 doi: 10.1073/pnas.1702155114

Murray, S., Holmes, J., Bellavia, G., Griffin, D., & Dolderman, D. (2002). Kindred spirits? the benefits of egocentrism in close relationships. *Journal of Personality and Social Psychology*, *82*(4), 563–581.

Newport, E. L., & Aslin, R. (2004a). Learning at a distance i. statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. Retrieved from https://doi.org/10.1016/S0010-0285(03)00128-2 doi: 10.1016/S0010-0285(03)00128-2

Newport, E. L., & Aslin, R. N. (2004b). Learning at a distance i. statistical learning of non-adjacent dependencies. *Cognitive psychology*, *48*(2), 127–162.

Nosofsky, R. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. Retrieved from https://doi.org/10.1037/0278-7393.14.4.700 doi: 10.1037/0278-7393.14.4.700

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Reis, H., & Shaver, P. (1988). Intimacy as an interpersonal process. In *Handbook of personal relationships: Theory, research and interventions* (p. 367–389). John Wiley & Sons.

Rinck, M., Glowalla, U., & Schneider, K. (1992). Mood-congruent and mood-incongruent learning. *Memory & Cognition*, *20*(1), 29–39. Retrieved from https://doi.org/10.3758/bf03208251 doi: 10.3758/bf03208251

Robbins, J., & Krueger, J. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, *9*(1), 32–47. Retrieved from https://doi.org/10.1207/s15327957pspr0901_3 doi: 10.1207/s15327957pspr0901_3

Rosch, E., Simpson, C., & Miller, R. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502. Retrieved from https://doi.org/10.1037/0096-1523.2.4.491 doi: 10.1037/0096-1523.2.4.491

Rosenbaum, M. (1986). The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, *51*(6), 1156–1166. Retrieved from https://doi.org/10.1037/0022-3514.51.6.1156 doi: 10.1037/0022-3514.51.6.1156

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, *30*(1), 87–104.

Russell, D. (1996). Ucla loneliness scale (version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, *66*(1), 20–40. Retrieved from https://doi.org/10.1207/s15327752jpa6601_2 doi: 10.1207/s15327752jpa6601_2

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, *76*(5), 805.

Saucier, G. (2009). Recurrent personality dimensions in inclusive lexical studies: Indications for a big six structure. *Journal of personality*, *77*(5), 1577–1614.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(04), 393–414. Retrieved from https://doi.org/10.1017/S0140525X12000660 doi: 10.1017/S0140525X12000660

Sened, H., Lavidor, M., Lazarus, G., Bar-Kalifa, E., Rafaeli, E., & Ickes, W. (2017). Empathic accuracy and relationship satisfaction: A meta-analytic review. *Journal of Family Psychology*, *31*(6), 742–752. Retrieved from https://doi.org/10.1037/fam0000320 doi: 10.1037/fam0000320

Skerry, A., & Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Current Biology*, *25*(15), 1945–1954. Retrieved from https://doi.org/10.1016/j.cub.2015.06.009 doi: 10.1016/j.cub.2015.06.009

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological review*, *96*(1), 58.

Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, *62*(5), 787–797. Retrieved from https://doi.org/10.1037/0022-3514.62.5.787 doi: 10.1037/0022-3514.62.5.787

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–103.

Tajfel, H. (2001). *Social stereotypes and social groups.* Psychology Press.

Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General*, *142*(1), 151–162. Retrieved from https://doi.org/10.1037/a0028232 doi: 10.1037/a0028232

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. Retrieved from https://doi.org/10.1016/j.tics.2017.12.005 doi: 10.1016/j.tics.2017.12.005

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences*, *113*(1), 194–199.

Terracciano, A., Abdel-Khalek, A. M., Adam, N., Adamovová, L., Ahn, C.-k., Ahn, H.-n., ... others (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science*, *310*(5745), 96–100.

Thornton, M. A., & Tamir, D. I. (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences*, *114*(23), 5982–5987. Retrieved from https://doi.org/10.1073/pnas.1616056114 doi: 10.1073/pnas.1616056114

Thornton, M. A., & Tamir, D. I. (2021). People accurately predict the transition probabilities between actions. *Science Advances*, *7*(9), eabd4995.

Tsai, J. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, *2*(3), 242–259. Retrieved from https://doi.org/10.1111/j.1745-6916.2007.00043.x doi: 10.1111/j.1745-6916.2007.00043.x

Vazire, S., & Mehl, M. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, *95*(5), 1202–1216. Retrieved from http://dx.doi.org.proxy.library.vanderbilt.edu/10.1037/a0013314

Vélez, N., & Gweon, H. (2019). Integrating incomplete information with imperfect advice. *Topics in cognitive science*, *11*(2), 299–315.

Waytz, A., & Mitchell, J. (2011). Two mechanisms for simulating other minds: Dissociations between mirroring and self-projection. In *Current directions in psychological science.* Retrieved from https://doi.org/10.1177/0963721411409007 doi: 10.1177/0963721411409007

Weisman, K., Johnson, M. V., & Shutts, K. (2015). Young children's automatic encoding of social categories. *Developmental science*, *18*(6), 1036–1043.

Welborn, B., & Lieberman, M. (2014). Person-specific theory of mind in medial pfc. *Journal of Cognitive Neuroscience*, *27*(1), 1–12. Retrieved from https://doi.org/10.1162/jocn_a_00700 doi: 10.1162/jocn_a_00700

Willis, J., & Todorov, A. (2006). First impressions making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. Retrieved from https://doi.org/10.1111/j.1467-9280.2006.01750.x doi: 10.1111/j.1467-9280.2006.01750.x

Wilson, T., & Gilbert, D. (2003). Affective forecasting. In *Advances in experimental social psychology* (Vol. 35, p. 345–411). Academic Press. Retrieved from https://doi.org/10.1016/S0065-2601(03)01006-2 doi: 10.1016/S0065-2601(03)01006-2

Wojcieszak, M. (2008). False consensus goes online: Impact of ideologically homogeneous groups on false consensus. *The Public Opinion Quarterly*, *72*(4), 781–791.

Woo, B., & Mitchell, J. (2020). Simulation: A strategy for mindreading similar but not dissimilar others? *Journal of Experimental Social Psychology*, *90*, 104000. Retrieved from https://doi.org/10.1016/j.jesp.2020.104000 doi: 10.1016/j.jesp.2020.104000

Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive science*, *42*(3), 850–884.

Wu, Y., Muentener, P., & Schulz, L. E. (2017). One-to four-year-olds connect diverse positive emotional vocalizations to their probable causes. *Proceedings of the National Academy of Sciences*, *114*(45), 11896–11901.

Zaki, J., & Ochsner, K. (2011). Reintegrating the study of accuracy into social cognition research. *Psychological Inquiry*, *22*(3), 159–182. Retrieved from https://doi.org/10.1080/1047840X.2011.551743 doi: 10.1080/1047840X.2011.551743

Zaki, J., & Williams, W. C. (2013). Interpersonal emotion regulation. *Emotion*, *13*(5), 803.

Zhao, Z., Thornton, M. A., & Tamir, D. I. (2020). Accurate emotion prediction in dyads and groups and its potential social benefits. *Emotion*. Retrieved from https://doi.org/10.1037/emo0000890 doi: 10.1037/emo0000890

A

Supplements to Chapter 1

A.1 State Selection

To select the states used in the current study, we applied k-medoids clustering to emotion transition probabilities from previous research (Thornton & Tamir, 2017).

In the referenced research, transition probabilities were derived from three large independent experience sampling datasets. In each dataset, participants were repeatedly prompted via text or phone app, at various intervals, to report on the emotional states they were experiencing.

Together, the three datasets provided 70642 reports by 10803 participants. Emotions that occur at two adjacent time points are counted as an emotion transition. These transitions were then tallied and normalized by frequency-based expectations to render three datasets of transitional probabilities.

For each of three independent datasets of transitional probabilities, we grouped the emotional states into five clusters via k-medoids, and the centroid state in each cluster was chosen. This selection process ensured that the emotional states used in our task spanned the state space, such that results would not be limited to certain sub-regions of the space.

A.2 Additional Person-Specificity Analysis

In both studies, we carried out an analysis of target-specificity that is conceptually similar to the permutation test included in the main text. In this analysis, we correlated each

participant's Self-Friend ratings with Self ratings from all other participants, except the participants friend. We thus obtained an empirical distribution of mismatched accuracy for each participant. We then calculated the percentile of each participant's correctly-matched accuracy within their own mismatched-accuracy distribution. The distribution of these percentile scores are shown in Figure S1.

In study 1, the distribution of percentile score is heavily negatively skewed, with a Median = 0.77 and Upper hinge = 0.97. Correlations between participants' person-specific ratings and the correctly matched targets' self ratings are systematically higher than those with all mismatched targets. This indicates that participants' person-specific ratings were indeed specific to their friend.

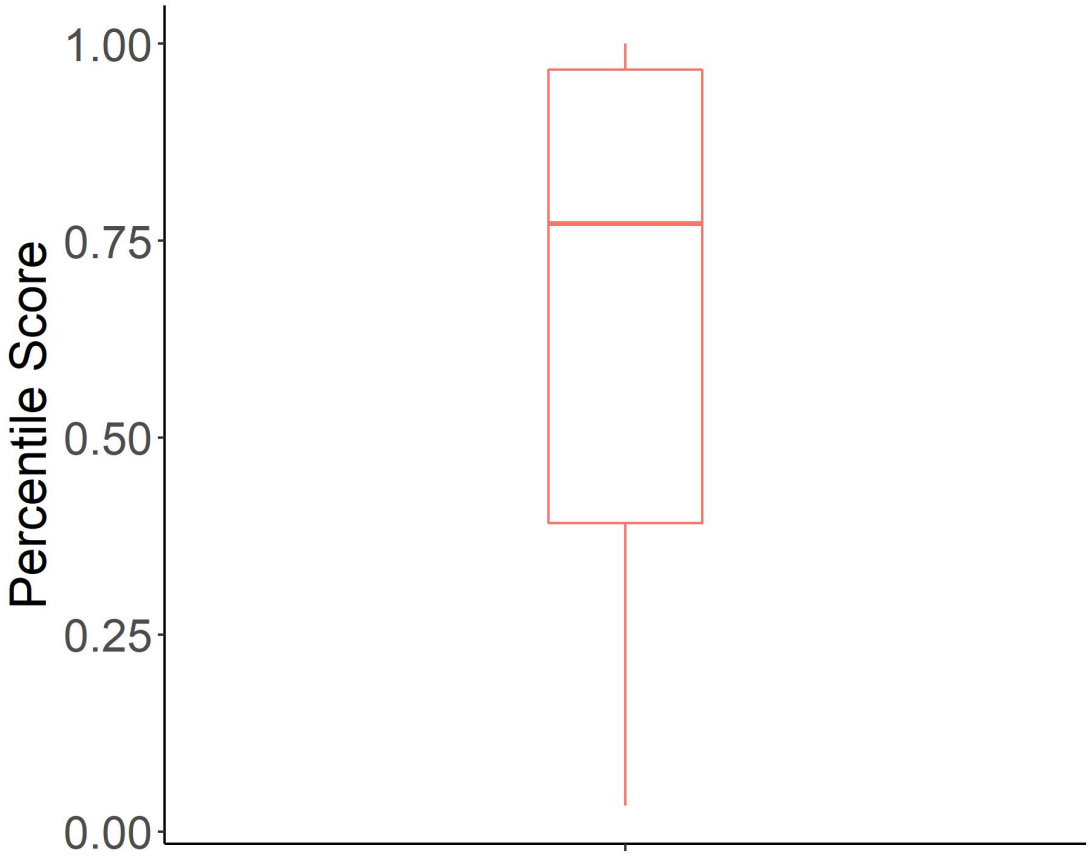

Figure A.1: Study 1: Percentile scores of correctly-matched person-specific accuracies in distributions of mismatched accuracies

In study 2, the distribution of percentile score is heavily negatively skewed, with a Me-

dian = 0.71 and Upper hinge = 0.97. Correlations between participants' person-specific ratings and the correctly matched targets' self ratings are systematically higher than those with all mismatched targets. This indicates that participants' person-specific ratings were indeed specific to their roommate.

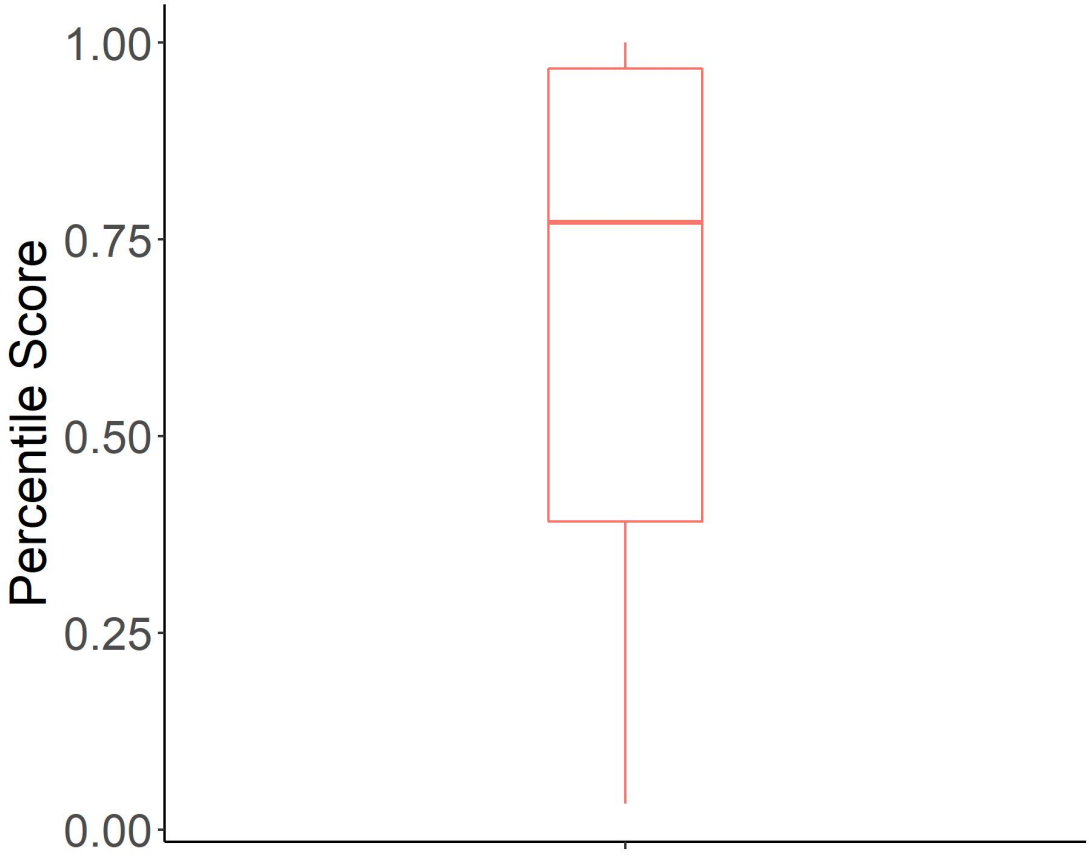

Figure A.2: Study 2: Percentile scores of correctly-matched person-specific accuracies in distributions of mismatched accuracies

A.3 Descriptive Statistics of the Social Success Measures

	Mean	Standard Deviation
Dyadic Success		
McGill Friendship Questionnaire - Respondent Affection	2.96	1.00
Perceived Closeness	85.89	13.89
Friendship Duration (months)	26.47	29.98
General Success		
Social Network Size	20.29	11.29
Social Support	6.16	3.80
Loneliness	42.41	8.85
Social Cognitive Ability		
Reading the Mind in the Eyes	27.47	3.52

Table A.1: Study 1: Means and Standard Deviations of Social Success Measures

	Mean	Standard Deviation
Dyadic Success		
McGill Friendship Questionnaire - Respondent Affection	1.63	1.43
Perceived Closeness (Scale of 10)	4.58	1.74
Reis Shaver Intimacy Model	48.39	12.68
General Success		
Social Network Size	16.39	10.36
Social Support	4.86	3.00
Loneliness	45.86	7.85
Days as group member	23.81	10.27
Social Cognitive Ability		
Reading the Mind in the Eyes	26.41	4.33
Relationship Duration	26.76	13.64
Days as group member	23.86	10.15

Table A.2: Study 2: Means and Standard Deviations of Social Success Measures.

A.4 Visualizations for the Accuracy-Success Correlations

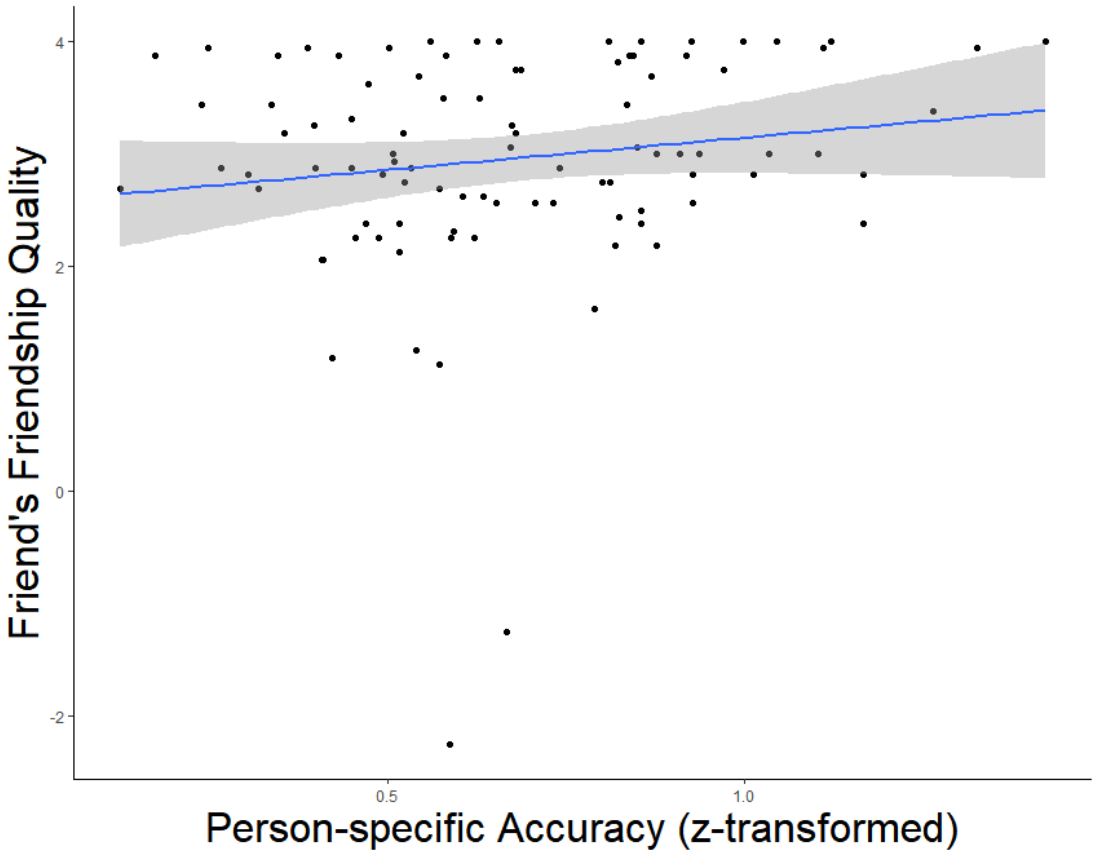

Figure A.3: Study 1: Correlation, person-specific accuracy and Friendship quality.

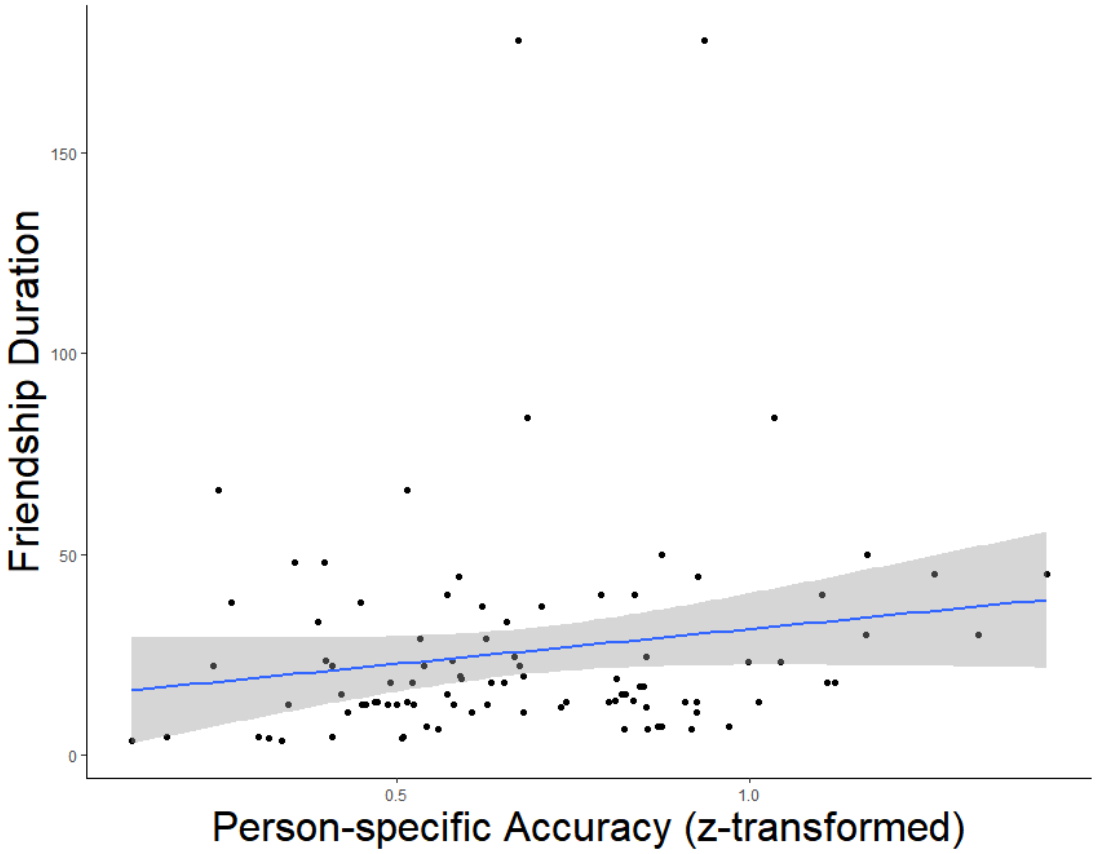

Figure A.4: Study 1: Correlation, person-specific accuracy and Friendship duration.

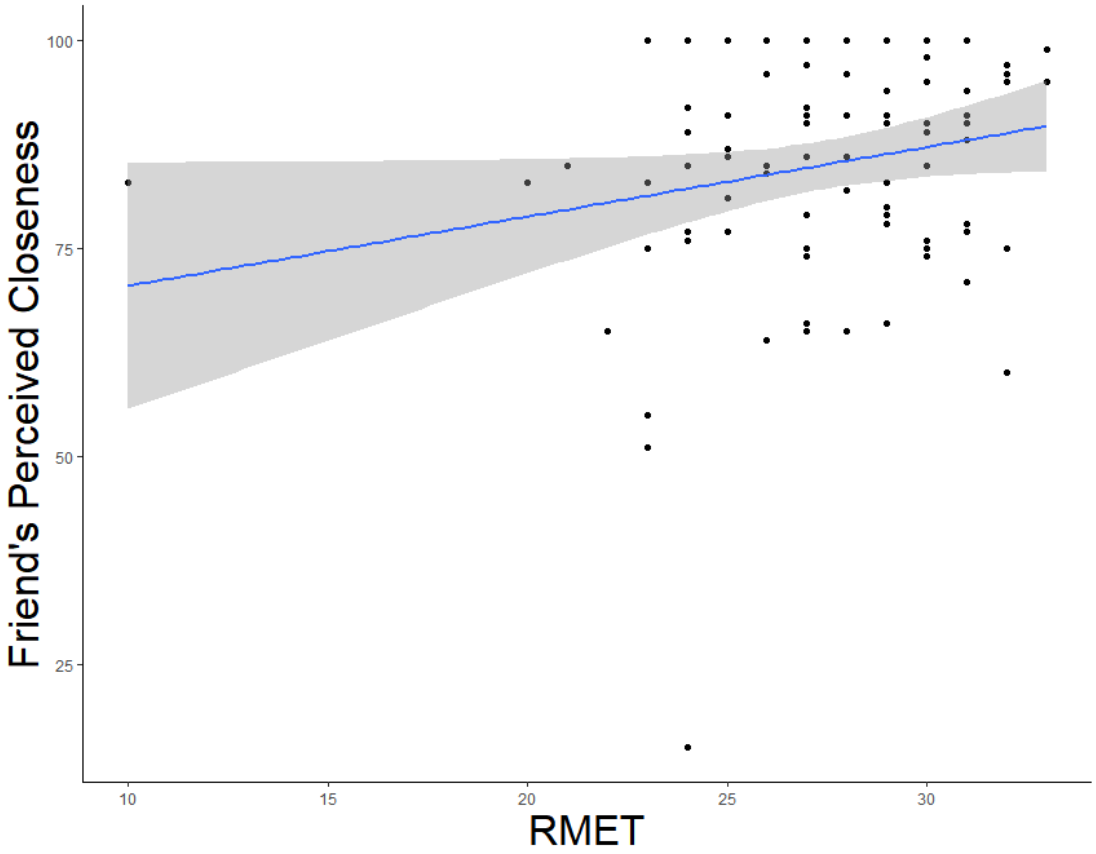

Figure A.5: Study 1: Correlation, RMET and Friend's perceived closeness

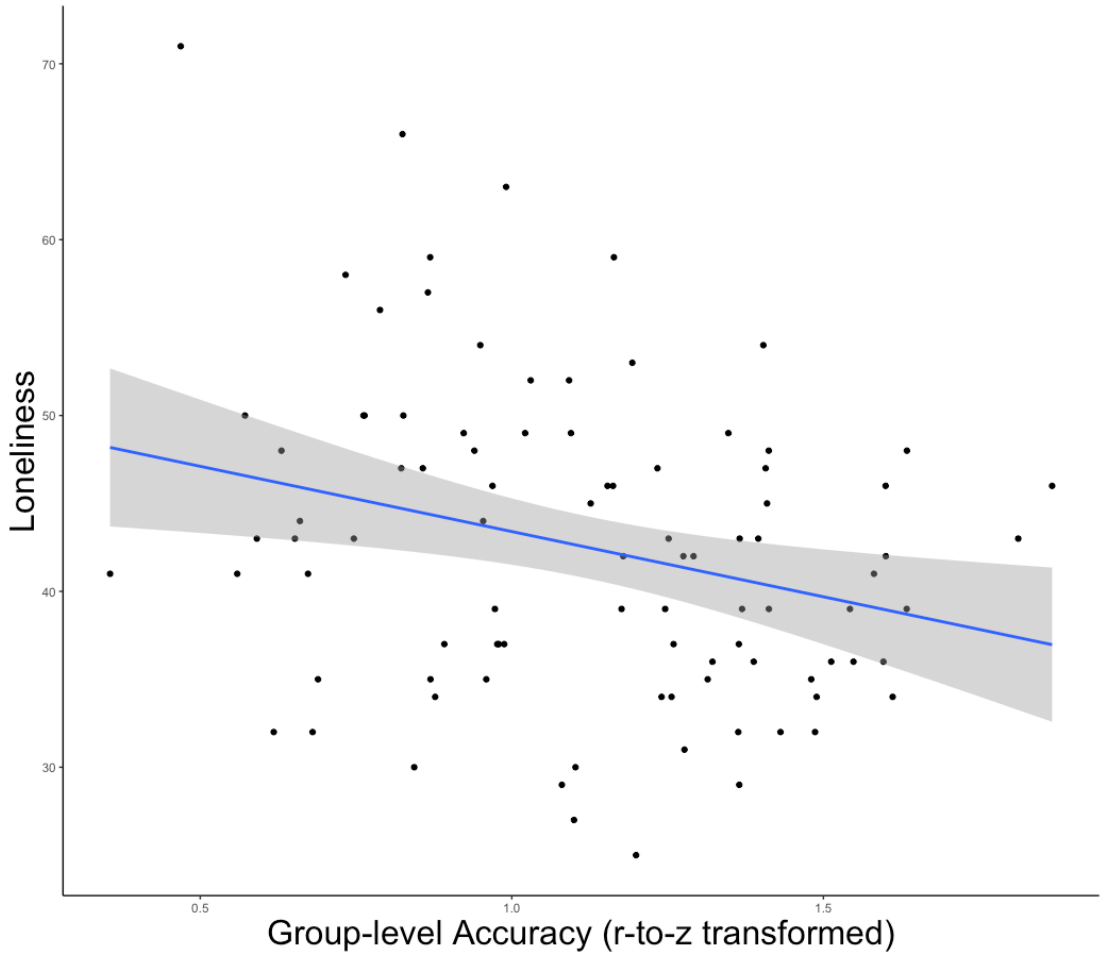

Figure A.6: Study 1: Correlation, Group-level accuracy and Loneliness

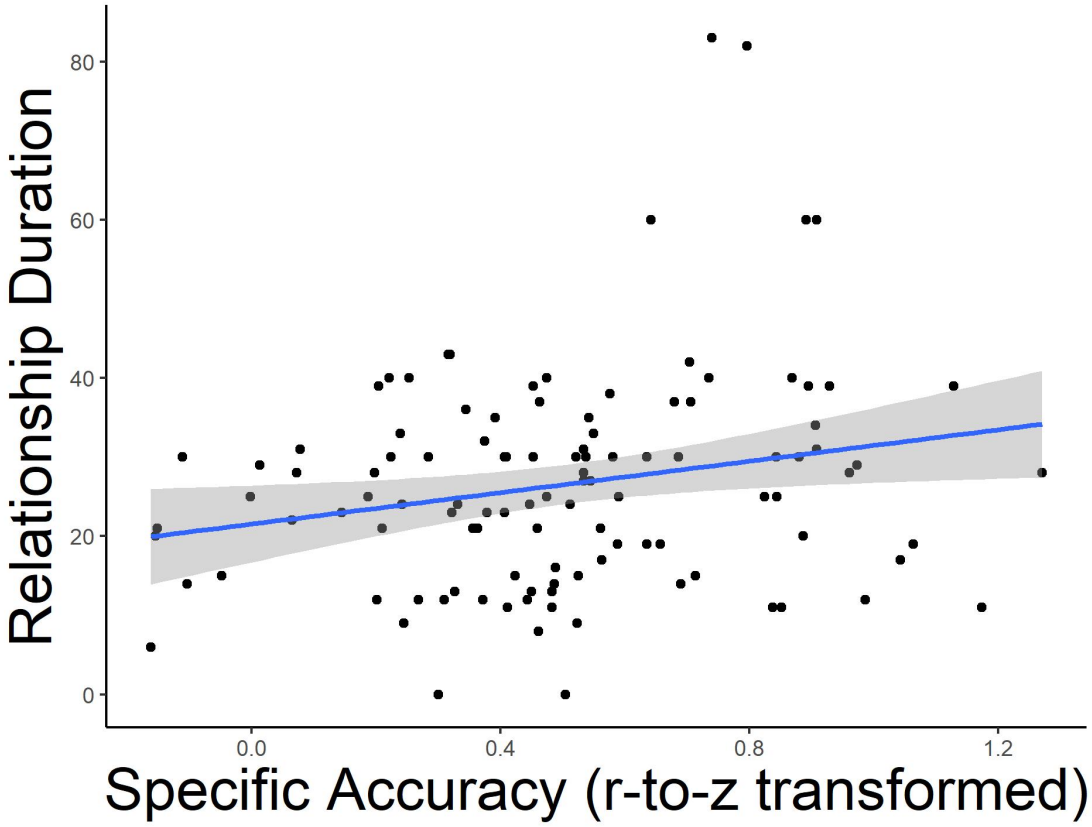

Figure A.7: Study 2: Correlation, Person-specific accuracy and Relationship Duration

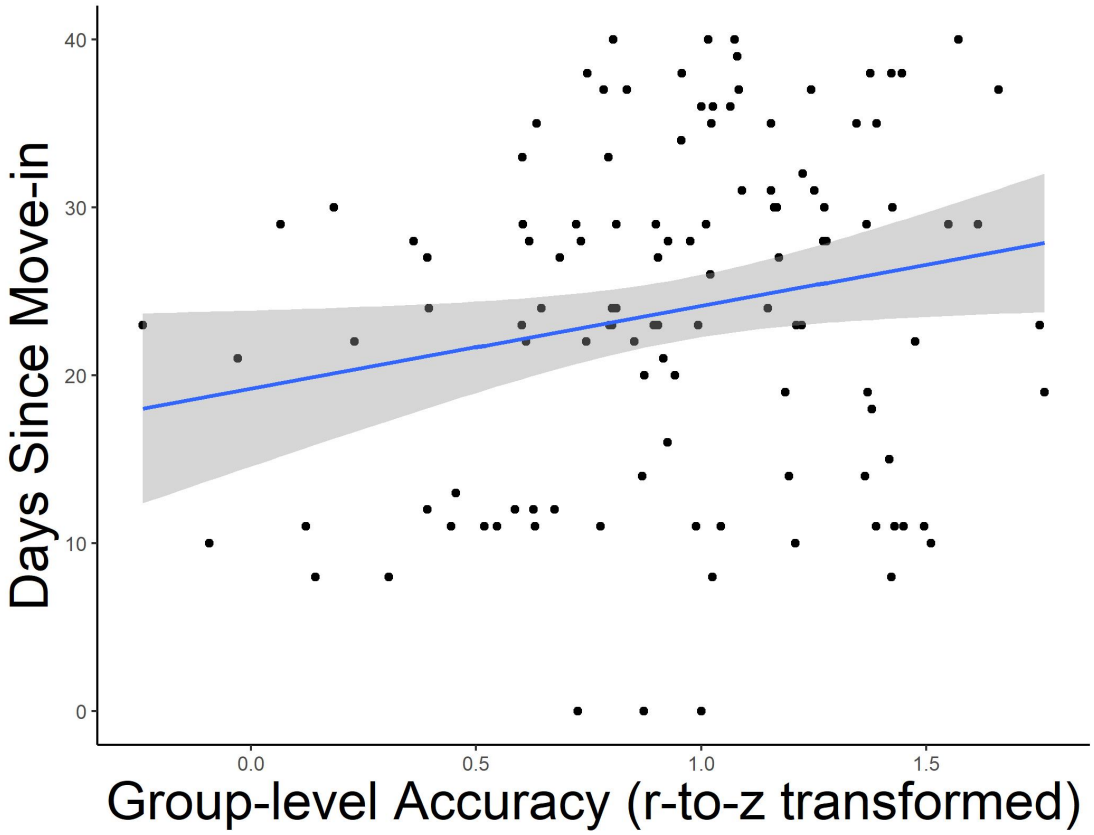

Figure A.8: Study 2: Correlation, Group-level accuracy and Days as a group member

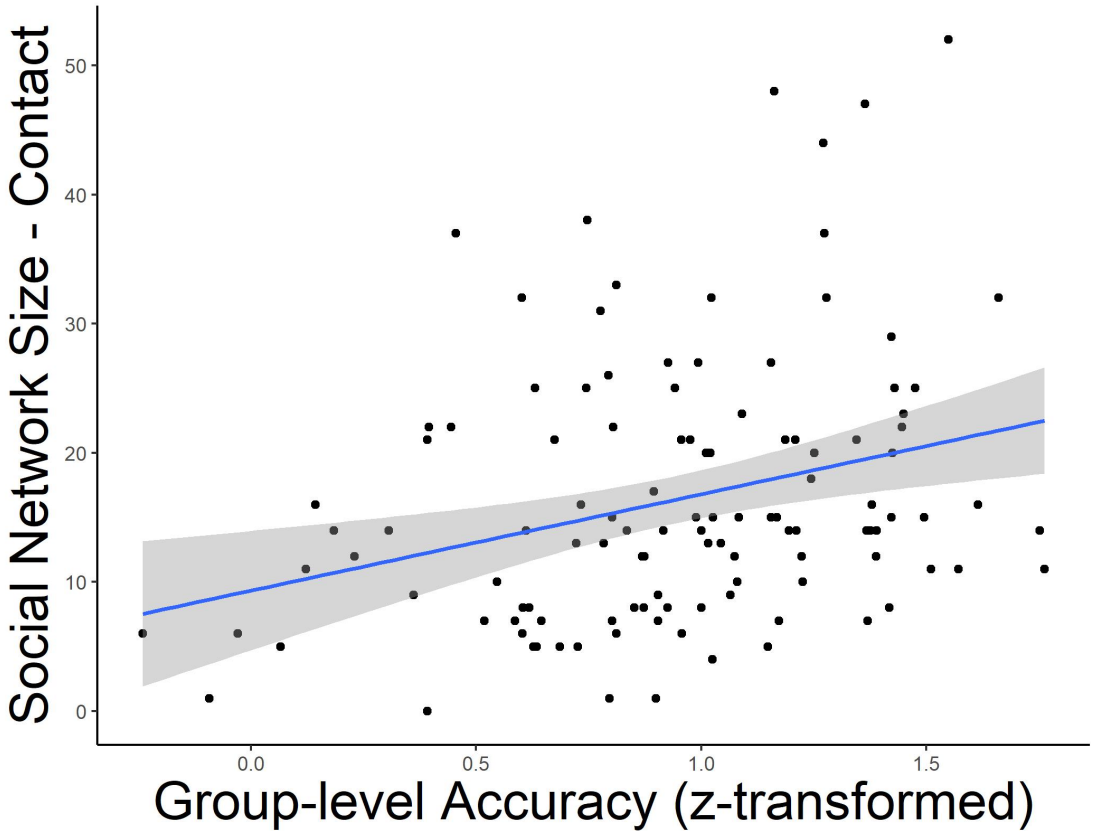

Figure A.9: Study 2: Correlation, Group-level accuracy and Social network size

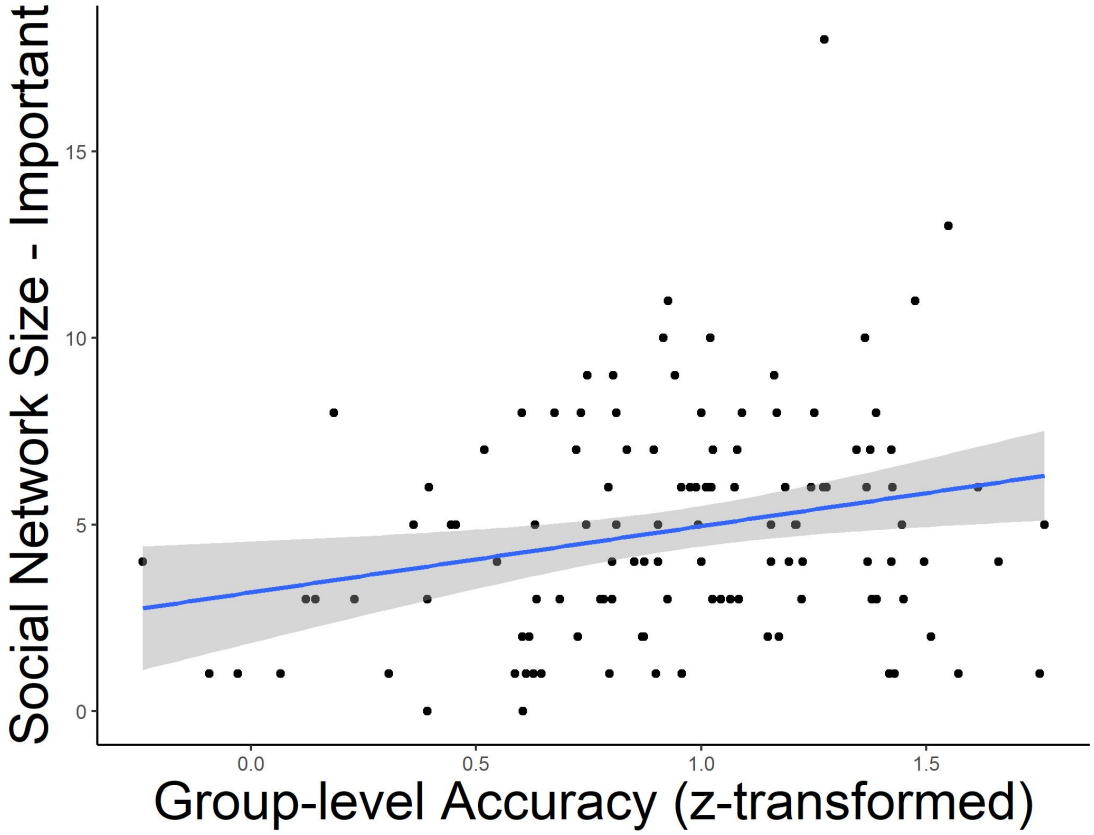

Figure A.10: Study 2: Correlation, Group-level accuracy and Support network size

A.5 Social Accuracy Model (SAM) based analyses

Componential approaches to interpersonal accuracy stress that accuracy in social perception can be further divided into different components such as normative accuracy, one ' s accuracy in perceiving the trait of an average target, and distinctive accuracy, one ' s accuracy in perceiving how a target differs from the average. These accuracy components conceptually map onto the effects of Group ratings and Friend ratings in the regression analysis in the main text. Despite the similarities, there are two key differences between the SAM and the current approach.

First, the SAM has most often been applied to data from paradigms where each perceiver rates multiple targets and each target is rated by multiple perceivers (e.g. round-robin). Resulting component estimates have been fruitfully linked to various important social phenomena such as well-adjustedness (Human, Carlson, Geukes, Nestler, & Back, 2020) and relationship development (Human & Biesanz, 2011). The dyad data modeled in the cur-

rent study, however, deviates from such a format substantially in that each subject rated and was rated by only one other subject in the sample. Second, the SAM treats target self ratings (d.g. Friend-actual ratings) as a validation measure used to model one's ratings for the target. The current approach, on the other hand, treats target self ratings as what's being modeled.

Nonetheless, due to conceptual similarity, we conducted additional multiple regression analyses following the social accuracy model (Biesanz, 2010) across both Study 1 and Study 2 in order to further examine person-specific accuracy.

Study1: To examine person-specific accuracy in the SAM-inspired model, we conducted a linear mixed effects regression where Friend ratings were included as the dependent variable; Friend-Actual, Self ratings, and Group ratings were included as predictors. Random intercepts were included for both subjects, nested in dyads, and for items (emotion pairs). Results from this analysis showed that Friend-Actual significantly predicted Friend ratings ($b = 0.09, \beta = 0.09, t(3478) = 8.56, p < .001$). Both Group-ratings ($b = 0.30, \beta = 0.31, t(6367) = 25.45, p < .001$) and Self ratings ($b = 0.35, \beta = 0.33, t(5458) = 26.98, p < .001$) also significantly predicted Friend ratings, reflecting known high collinearity. Study 2: We conducted the same model analyses as Study 1, and found that Friend-Actual significantly predicted Friend ratings ($b = 0.06, \beta = 0.06, t(7397) = 5.93, p < .001$). Both Group-ratings ($b = 0.30, \beta = 0.30, t(8406) = 26.53, p < .001$) and Self ratings ($b = 0.26, \beta = 0.27, t(8538) = 24.67, p < .001$) also significantly predicted Friend ratings. These results are consistent with the primary analyses in the main manuscript.

A.6 Additional Exploratory Analyses

Association between accuracy and social success across the specific--general boundary. In the primary analyses, we tested the *a priori* hypotheses that person-specific accuracy will be associated with relationship success, and group-level accuracy will be associated with general social success. However, it is also possible that these associations exist across the specific-general boundary. To the extent people who are more accurate about specific individuals are also more accurate at the group level, we might expect to see a positive relation between group-level accuracy and relationship success, and vice versa. It is also possible that person-specific accuracy uniquely predicts relationship success, and group-level accuracy general social success. Here we explore these possibilities.

We first examined the extent to which person-specific accuracy and group-level accuracy are related. A correlation analysis reveals that the two are strongly but not perfectly associated in both Study 1 ($r = .50$, 95% bootstrapped $CI = [.33, .65]$) and Study 2 ($r = .60$, 95% bootstrapped $CI = [.46, .70]$), indicating that the two might share common factors,

but are not identical.

We next examined whether relationship success is uniquely predicted by person-specific accuracy. To do so, we used person-specific and group-level accuracy to simultaneously predict each of the relationship success measures. We did so using linear mixed effects models, with random intercepts at the dyad level. Fixed effect estimates from these models are reported in the table below. We found ambiguous evidence in Study 1 that person-specific accuracy predicted friend's relationship satisfaction as measured by MFQRA after controlling for group level accuracy ($p = .054$). We found no other association between either person-specific accuracy or group-level accuracy and relationship success across the models tested.

	Person-Specific	Group-level
Study 1.		
McGill Friendship Questionnaire - Respondent Affection	.24	-.08
Perceived Closeness (Scale of 10)	.17	-.06
Study 2.		
McGill Friendship Questionnaire - Respondent Affection	.08	.00
Perceived Closeness (Scale of 10)	.02	-.08
Reis Shaver Intimacy Model	.13	-.02

Table A.3: Standardized betas, linear mixed effects model simultaneously predicting relationship success with person-specific accuracy and group-level accuracy

We also examined whether relationship success is uniquely predicted by person-specific accuracy. We used person-specific and group-level accuracy to simultaneously predict each of the general social success measures in linear mixed effects models, with random intercepts at the dyad level. Fixed effect estimates from these models are reported in the table below. We found ambiguous evidence in Study 1 that person-specific accuracy predicted friend's relationship satisfaction as measured by MFQRA after controlling for group level accuracy ($b = .24, p = .054$). We found no other association between either person-specific accuracy or group-level accuracy and relationship success across the models tested. We found in Study 1 that group-level accuracy negatively predicted loneliness after controlling for person-specific accuracy ($p < .01$). In Study 2, we found that after controlling for person-specific accuracy, group-level accuracy positively predicted social network size ($p < .01$) and support network size ($p < .01$).

Overall, we found some evidence that, under high collinearity, group-level accuracy uniquely predicted general social success after controlling for person-specific accuracy, but no evidence that person-specific accuracy uniquely predicted relationship success.

	Person-Specific	Group-level
Study 1.		
Loneliness	.05	-.30**
Social network size	.06	-.19
Support network size	.07	-.02
Study 2.		
Loneliness	.01	-.15
Social network size	-.09	.35**
Support network size	-.04	.29**

Table A.4: Standardized betas, linear mixed effects model simultaneously predicting general social success with person-specific accuracy and group-level accuracy

Accuracy-success link quickly emerges early in relationship. In Study 2, we found no significant relation between accuracy measures and corresponding social success measures among new college roommates, though many of the relations were numerically in the right direction. One possibility we did not consider was that the accuracy-success link might be moderated by relationship duration and group membership duration. Much in the same way levels of accuracy increased over the short period of early relationship development, the association between accuracy and social success might also emerge quickly. Here we explore such a possibility.

We first tested whether those new college roommates who completed the study later on showed stronger links both between person-specific accuracy and relationship success. We did so using linear mixed effects regressions to model the interaction between relationship duration and person-specific accuracy in predicting relationship success. We found a significant positive interaction between the two in predicting the partner's relationship satisfaction ($b = 0.10, \beta = 0.28, t(111.7) = 2.69, p < .01$). For roommates who had known each other longer, one's person-specific accuracy better predicted their roommate's relationship quality ratings. We found a comparable interaction effect predicting the partner's response on the Reis-Shaver Intimacy scale ($b = 1.03, \beta = 0.31, t(112.0) = 3.20, p < .01$), but not the partner's perceived closeness ($b = 0.05, \beta = 0.10, t(110.0) = 0.99, p = .32$).

We then tested whether those new college students who completed the study later on showed stronger links between group-level accuracy and relationship success. We did so using linear mixed effects regressions, modeling the interaction between group membership duration and group-level accuracy in predicting general social success. We did not, however, observe any significant interaction effect across models predicting loneliness ($b = -0.24, \beta = 0.12, t(110.4) = -1.25, p = .22$), social network size ($b = 0.08, \beta = 0.03, t(107.1) = 0.32, p = .75$), or support network size ($b = -.01, \beta = -.02, t(111.4) =$

$-0.19, p = .85$). Overall, we found preliminary evidence that the link between person-specific accuracy and relationship success might emerge relatively quickly during relationship development. Future research should further investigate this finding.

B

Supplements to Chapter 2

B.1 Simulations for the Split-bootstrap Procedure

B.1.1 Split-bootstrap removes spurious construct correlations

In our primary analyses, we examined the relations among constructs including egocentrism, accuracy, typicality, and dyadic similarity. To score these constructs for each participant, we first computed bivariate correlations between different variables obtained from the emotion transition task. The resulting correlation coefficients were than Fisher's z-transformed and used in further analyses. As mentioned in the main manuscript, a potential challenge for this approach was that the individual level correlation scores of different constructs can share underlying covariates. For example, we might want to examine the relation between egocentrism and accuracy, but the two share an underlying covariate. Egocentrism scores were calculated by correlating each participant's Self ratings with their Target ratings, whereas accuracy scores were calculated by correlation each participant's Target ratings with their target's Self ratings. The two shared participant Target ratings as a low-level covariate.

Using simulation, we concretely demonstrated the potential problem as a result of this covariate sharing. To simulate our analytical scenario, we assumed that there were as many participants as in the actual data from Study 3. In each iteration, we generated three random variables A, B, and C for each participant. These variables corresponded to the low-level covariates in our analyses. The construct scores were calculated by correlating r_{AB} and r_{AC} at the participant level, the target estimate is the correlation between these two coeffi-

cients across all participants, and B is the shared covariates between the construct scores. Across iterations, we varied the a priori correlation between A and C from 0 to 1 at .005 intervals, whereas r_{AB} and r_{AC} had no explicit constraint. There were 10 iterations at each step of r_{AC} value, resulting in total of 2000 iterations. The expected construct correlation (correlation between r_{AB} and r_{AC}) increased linearly as a function r_{AC}, whereas the variability of the construct score decreased. This pattern suggests that, in our analytical scenario, sharing correlated low-level covariates can introduce spurious correlation between construct scores (Figure B.1).

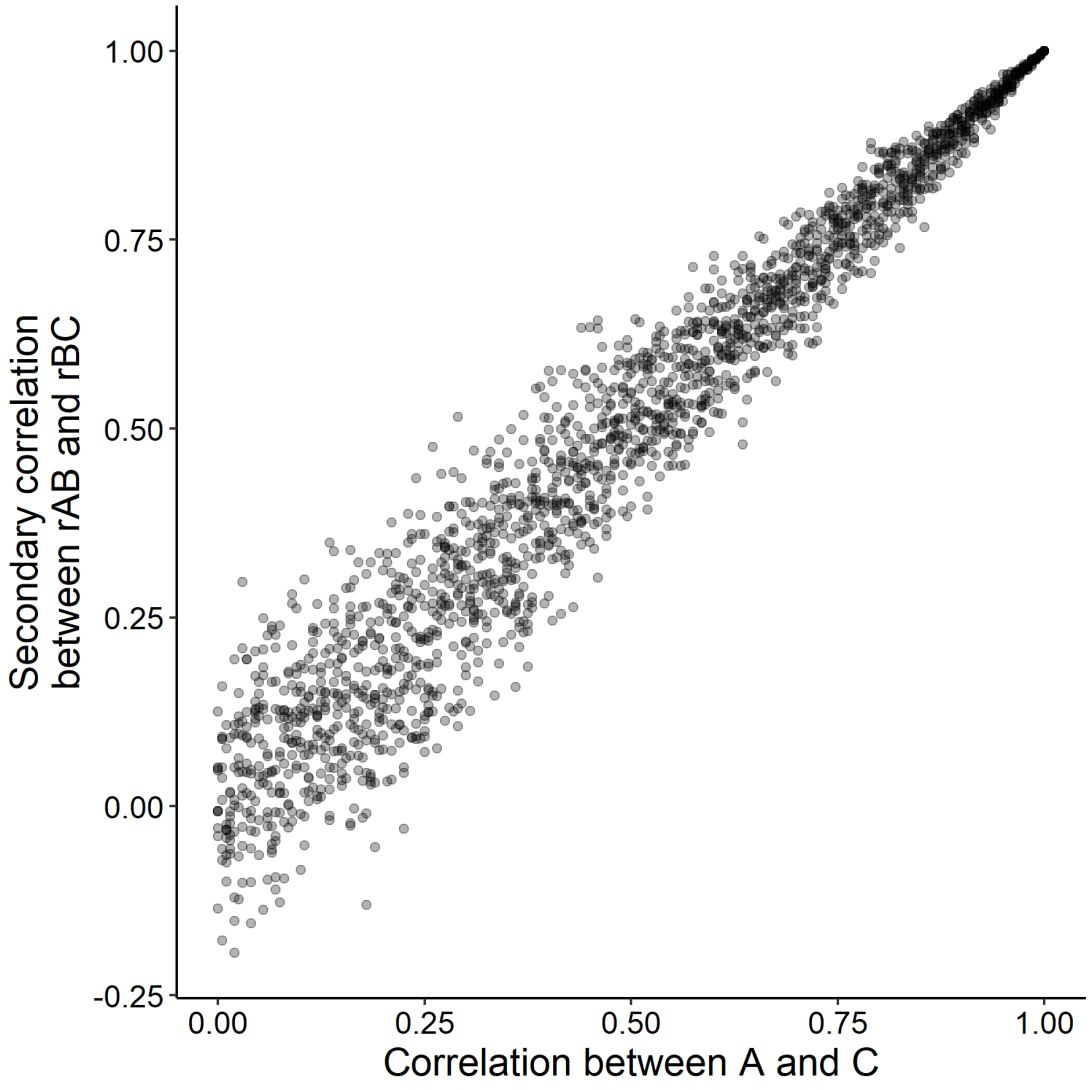

Figure B.1: Covariate sharing introduces spurious construct correlations. In an iterative simulation, the correlations between two constructs that shared a covariate (correlation between r_{AB} and r_{AC}; y-axis) linearly tracked the correlation between the unshared low level covariates (r_{AC}; x-axis).

To address this issue, we applied an iterative subset-and-bootstrap procedure to eliminate the spurious construct correlation produced by covariate sharing. In each iteration of this procedure, we first divided the 75 state pairs from the emotion transition task into two equally sized subsets. Each subset was used to estimate one individual level score. For example, to estimate the relation between egocentrism and accuracy, we first divided the

75 items into two sets (a 37-38 split in a single iteration, but equally sized in the long run). For each participant, we obtained their egocentrism scores by correlating their Self ratings and Target ratings on item set 1, and then applying Fisher's z-transformation. We obtained accuracy scores for the same participant by correlating their Self ratings and their target's Self ratings on item set 2, and then applying Fisher's z-transformation. Once these scores were computed, we bootstrapped over participants by repeatedly resampling, with replacement, the same number of participants as the actual sample. For each sample, we calculated the correlation between egocentrism an accuracy scores. These two steps were repeated in a nested fashion to obtain an empirical distribution of the relation of interest.

We demonstrate the efficacy of this procedure using the same simulated data. In each iteration of the simulation procedure, we first divided A, B, and C into equally sized subsets. Half of A and B were used to compute r_{AB}, whereas the unused half of B and the corresponding half of C were used to compute r_{BC}. These two correlation coefficients were then used to calculate the construct correlation. When the low level covariates were split, the construct correlation no longer tracked the correlation between A and C. This suggests that the split-and-bootstrap procedure effectively removed the spurious construct correlation introduced by covariate sharing (Figure B.2).

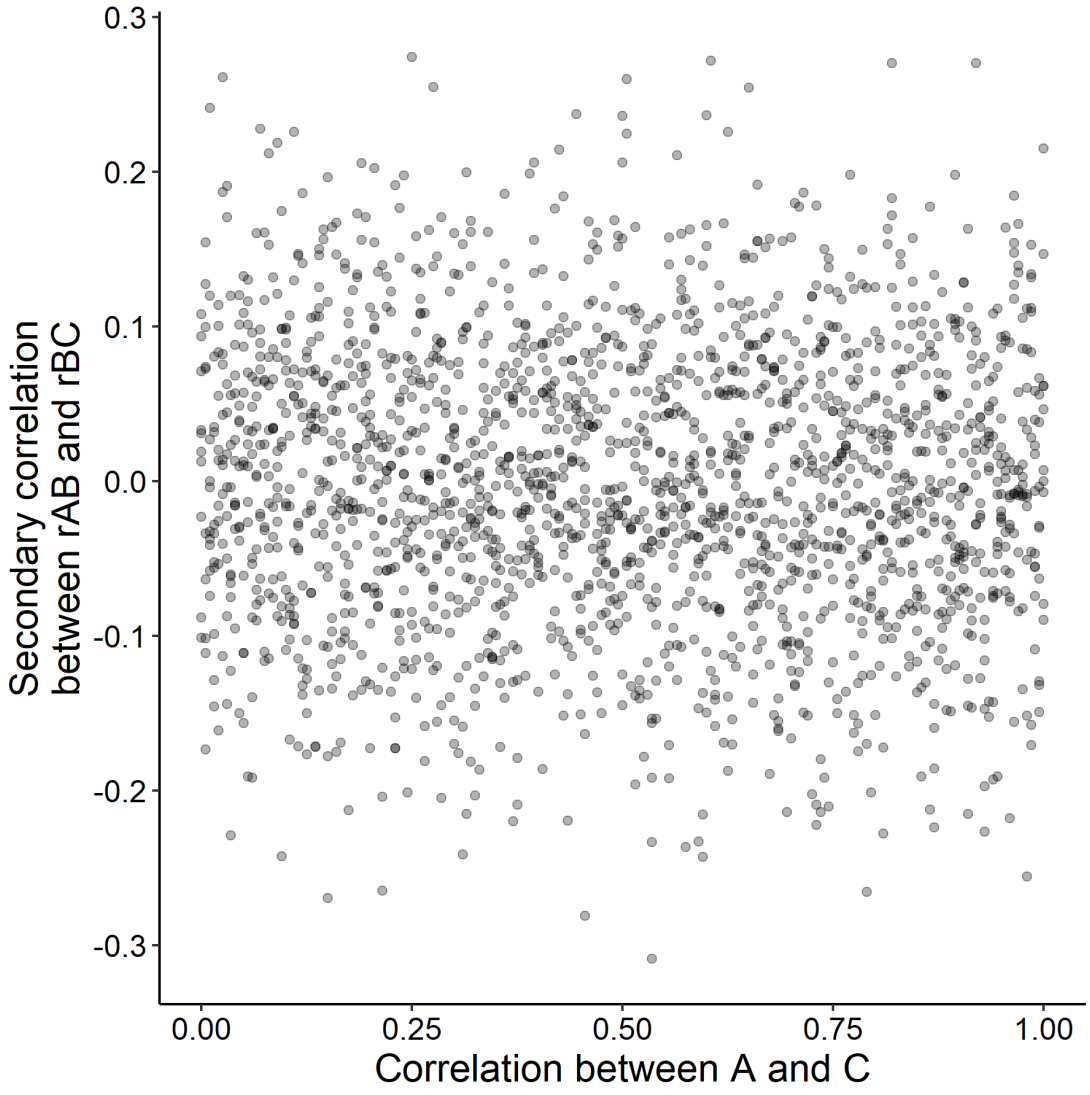

Figure B.2: Split-bootstrap removes spurious construct correlations from covariate sharing. When the simulated data were analyzed using the split-bootstrap procedure, the construct correlations was no longer related to the correlations between unshared covariates.

B.1.2 Split-bootstrap preserves true construct correlations

This previous simulation procedure took a bottom-up approach, specifying *a piori* correlations between low level covariates, and demonstrating that the split-bootstrap procedure can remove spurious construct correlations introduced by covariate sharing. However, this approach was agnostic of the true construct correlations. If there were genuine correlations

between the construct scores, would they preserved by the split-bootstrap procedure?

To answer this question, we conducted a second, top-down simulation. We start by specifying the true construct correlation ($r_{r_{AB},r_{AC}}$)) at different levels. We used this a priori construct correlation to randomly generate construct scores for N participants. For each participant, we then used these generated r_{AB} and r_{AC} scores to generate low level covariates A, B, and C at varying levels of r_{AC}. Finally, we subjected these top-down generated covariates to the split-bootstrap procedure.

The recovered construct correlations tracked with the a priori construct correlations but not the rAC (Figure B.3). This suggested that the split-bootstrap procedure can indeed preserve the true construct correlations and remove spurious ones introduced by covariate sharing

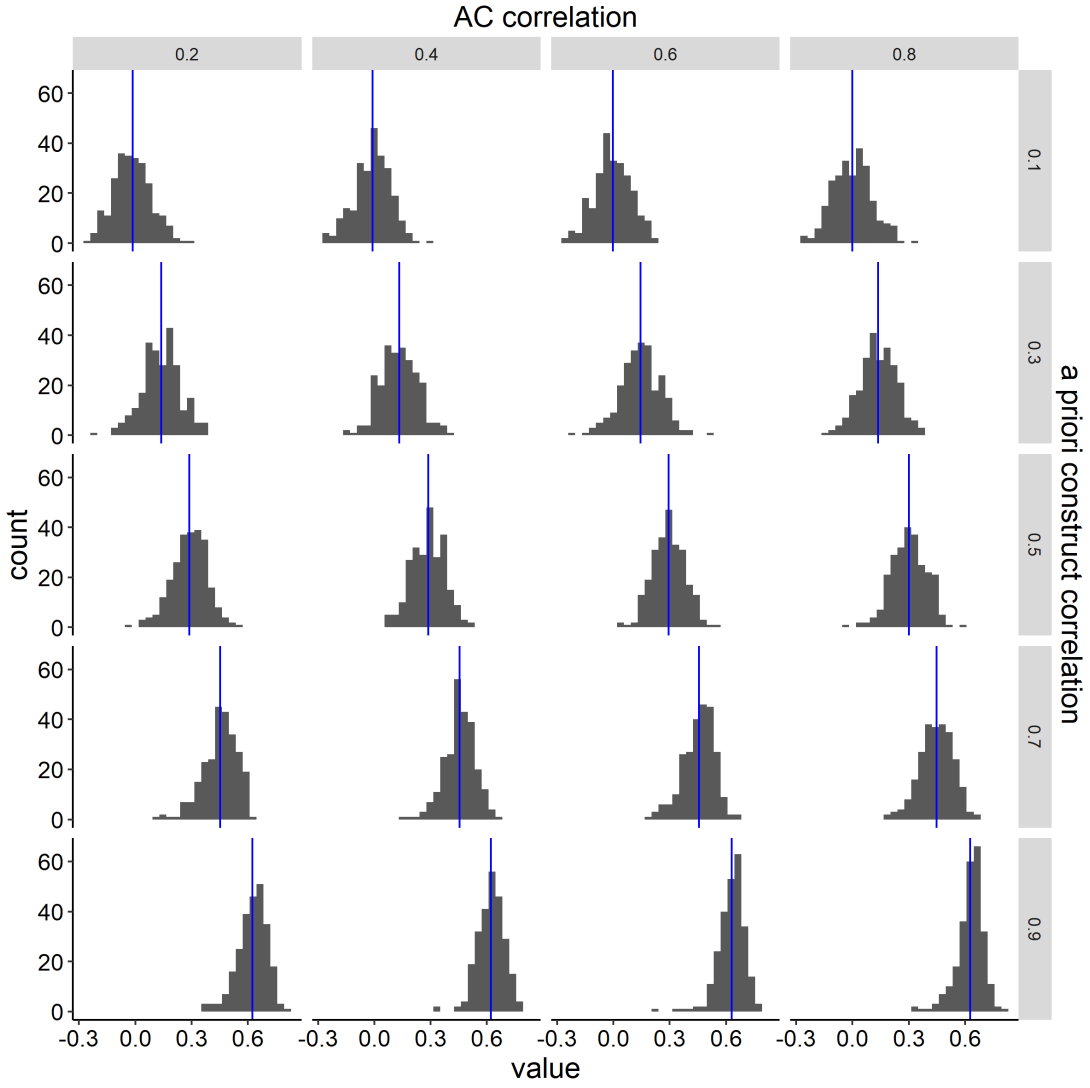

Figure B.3: Split-bootstrap recovers the true construct correlations. Distributions of estimated construct correlations as a function of both the true construct correlations (grid y-axis) and the a prior correlations between A and C (grid x-axis). The expected values of the estimated construct correlation increased as the true construct correlations increase, but was not affected by changes in the *a priori* A-C correlations.

B.2 Dyadic Similarity and Egocentrism Independently Predicts Accuracy

In addition to the association between egocentrism and accuracy reported in the manuscript, we also identified a positive relation between accuracy and dyadic similarity. In Study 3, the

accuracy of participants' emotion prediction was positively associated with both domain-specific ($r(114) = .69, 95\%CI = [.55, .79]$) and domain-general ($r(114) = .22, 95\%CI = [.04, .39]$) measures of dyadic similarity (Figure B.4). We replicated these relations in Study 4, where accuracy was positively associated with domain-specific ($r(114).72, 95\%CI = [.60, .81]$) and domain general ($r(114) = .44, 95\%CI = [.26, .59]$) similarity (Figure B.4).

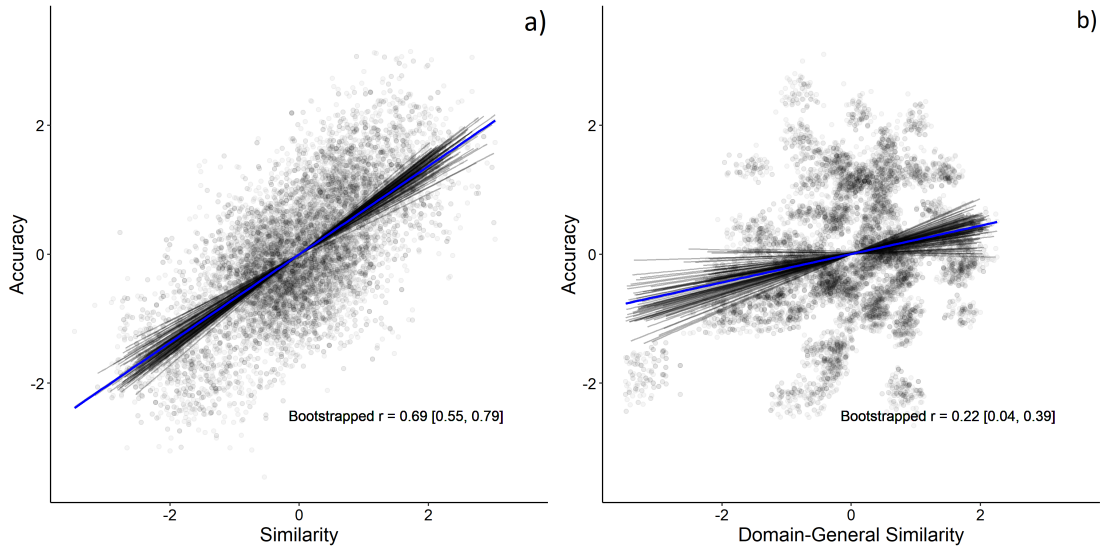

Figure B.4: Correlation between Accuracy and Dyadic Similarity scores. People who were more similar to their roommate, both in terms of their (a) emotion dynamics and (b) personality, more accurately predicted their partner's emotion transitions. Points represent 100 bootstrap samples. Each black line represents the best fitting regression line for a sample. Blue line represents the mean bootstrapped correlation.

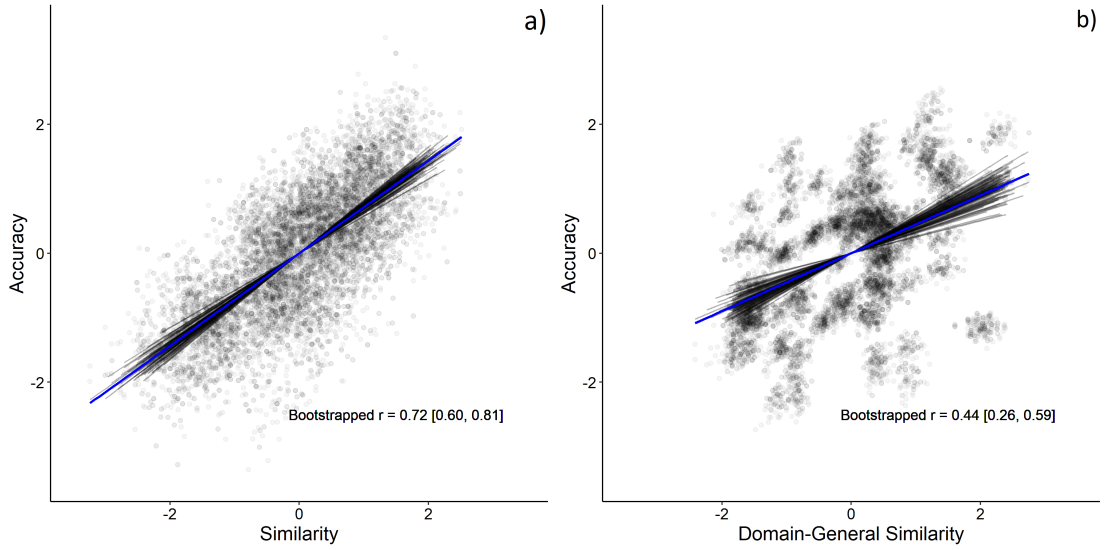

Figure B.5: Correlation between Accuracy and Dyadic Similarity scores in the replication sample. People who were more similar to their roommate, both in terms of their (a) emotion dynamics and (b) personality, more accurately predicted their partner's emotion transitions. Points represent 100 bootstrap samples. Each black line represents the best fitting regression line for a sample. Blue line represents the mean bootstrapped correlation.

Furthermore, we probed whether egocentrism was particularly effective when the rater and target were more similar. That is, the association between egocentrism and prediction accuracy should be positively moderated by dyadic similarity. We conducted a linear regression testing the interaction between egocentrism scores and similarity scores in predicting accuracy scores. If egocentrism works better the more similar the rater and the target, we expect to see a positive interaction between the two predictors.

In Study 3, we did not find evidence to support the hypothesized interaction ($b = 0.01, 95\%CI = [-.19, .22]$). However, we found that egocentrism ($b = 0.25, 95\%CI = [.04, .46]$) and similarity ($b = 0.51, 95\%CI = [.31, .69]$) both uniquely predicted accuracy after controlling for each other. We found a similar patter of results in Study 4. There was no evidence for the hypothesized interaction ($b = 0.06, 95\%CI = [-.13, .24]$). However, we replicated the the independent effects of egocentrism ($b = 0.23, 95\%CI = [.01, .43]$) and similarity ($b = 0.55, 95\%CI = [.35, .72]$).

These findings reflected two phenomena. On one hand, in a naturalistic college community, people might be largely similar, such that taking a more egocentric approach yielded more accurate predictions independent of the gradations in dyadic similarity. On the other hand, people overall engaged in a sufficient level of egocentrism such that those who were more similar to their target were able to achieve greater accuracy.

B.3 Dyadic Similarity Mediates the Typicality-Accuracy Link

B.3.1 Study 3

Given that dyadic similarity was associated with accuracy independently of how much one engages in egocentrism, we further evaluated the possibility that similarity also independently mediates the link between typicality and prediction accuracy. That is, holding the level of egocentrism constant, more typical perceivers tend to also be more similar to their target, and thus achieve greater accuracy. In a mediation analysis, we tested whether similarity scores accounted for the association between perceiver typicality scores and accuracy scores. If more typical individuals achieve greater accuracy by virtue of being more similar to their target, we expect that similarity should mediate the typicality-accuracy link.

The total effect using typicality to predict accuracy was $b = .44\,(95\%CI = [.26, .60])$. We detected a positive indirect effect through similarity ($b = .32, 95\%CI = [.18, .46]$). After mediation, the effect of typicality was reduced ($b = .12, 95\%CI = [-.08, .29]$), indicating that the association between typicality and accuracy was mediated by dyadic similarity. However, this finding did not extend to the domain general analyses. The total effect using domain general typicality to predict accuracy was b = .22 ($95\%CI = [.003, .41]$), and the direct effect after mediation by similarity was reduced to b = .15 ($95\%CI = [-.12, .42]$). While the indirect effect through domain general similarity was positive, its confidence interval did not exclude 0 ($b = .07, 95\%CI = [-.09, .24]$). This analysis thus did not provide conclusive evidence that the association between domain general typicality and accuracy was mediated through domain general similarity (Figure B.6).

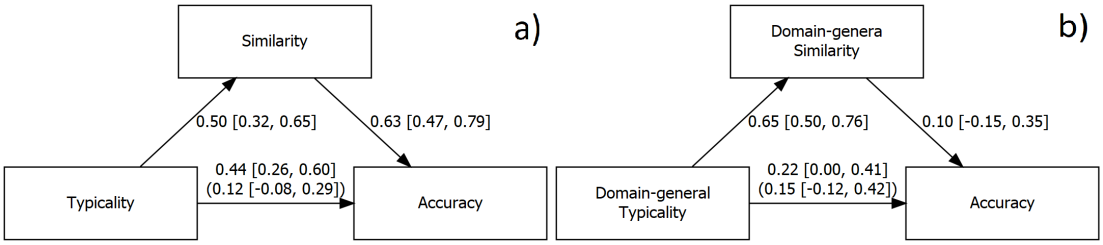

Figure B.6: Dyadic similarity mediated the Typicality-Accuracy link. Typical perceivers more accurately predicted novel targets' emotion transitions, and they did so by virtue of being more similar to their target.

Finally, we tested the hypothesis that using more egocentrism, together with being more similar to a target independently enable more typical individuals to achieve greater accuracy in their emotion predictions. In a multiple mediation model, we tested whether egocentrism and similarity simultaneously mediated the link between typicality scores and accu-

racy scores. We found evidence in support of simultaneous mediation. The total effect, using typicality scores to predict accuracy scores was $b = .39$ ($95\%CI = [.20, .57]$). We detected a positive indirect effect through egocentrism ($b = .19, 95\%CI = [.05, .34]$), as well as a positive indirect effect through similarity ($b = .22, 95\%CI = [.11, .36]$). The direct effect of typicality, after the effects of both egocentrism and similarity were accounted for, was reduced to -.02 ($95\%CI = [-.24, .20]$). This suggests that association between typicality and accuracy was simultaneously mediated through both egocentrism and dyadic similarity. Results from the domain general analyses were consistent with the aforementioned single mediation models. We found a mediating effect for egocentrism, but not for similarity. The total effect of using domain general typicality to predict accuracy was $b = .22$ ($95\%CI = [.01, .41]$). The indirect effect mediated through domain general similarity was positive but the 95% confidence interval did not exclude zero ($b = .06, 95\%CI = [-.09, .21]$). On the other hand, there was a positive indirect effect through egocentrism ($b = .12, 95\%CI = [.02, .21]$). The effect of domain general typicality was reduced after mediation, and the direct effect was $b = .04$ ($95\%CI = [-.28.19]$). These findings robustly supported the idea that egocentrism mediates the typicality-accuracy association, and provided partial support for the mediating effect of dyadic similarity (Figure B.7).

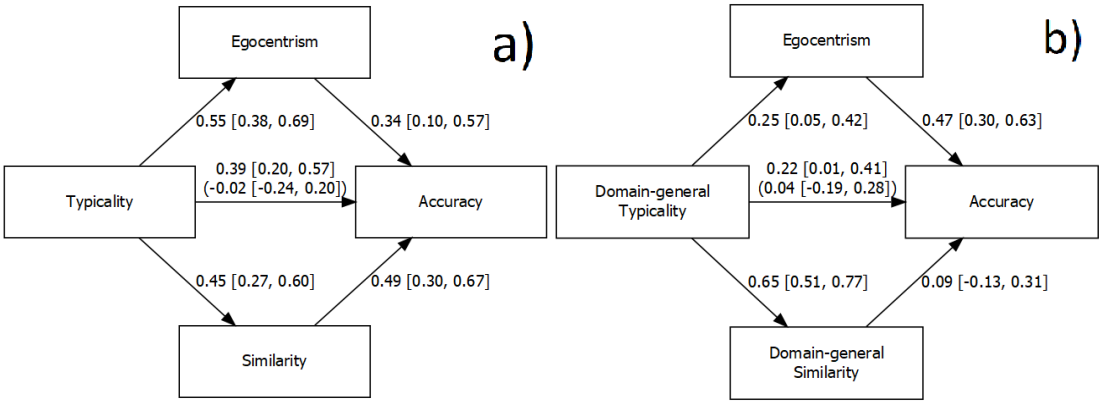

Figure B.7: Dyadic similarity and egocentric inference both independently mediated the Typicality-Accuracy link.

B.3.2 Study 4

We replicated the mediation analyses involving dyadic similarity using data from Study 4. We first tested whether a person's typicality might affect accuracy by influencing how similar they would be to a novel target. In a mediation model, the association between perceiver

typicality scores and accuracy scores was partially mediated by dyadic similarity scores. The total effect of typicality predicting accuracy was $b = .47$ ($95\%CI = [.27, .63]$). We detected a positive indirect effect through similarity ($b = .36, 95\%CI = [.23, .52]$). After mediation, the effect of typicality was reduced ($b = .11, 95\%CI = [-.11, .31]$), indicating that the association between typicality and accuracy was mediated by dyadic similarity. Unlike Study 3, the mediating effect of similarity in Study 4 extended into domain general analyses. That is, domain general dyadic similarity partially mediated the association between domain general typicality and accuracy. The total effect of domain general typicality predicting accuracy was $b = .34$ ($95\%CI = [.16, .50]$). We detected a positive indirect effect through similarity ($b = .12(95\%CI = [.03, .22]$). After mediation, the effect of domain general typicality was reduced ($b = .22, 95\%CI = [.03, .40]$). These replicated results further supported that more typical individuals can make more accurate predictions about novel targets by virtue of being more similar to them (Figure S10).

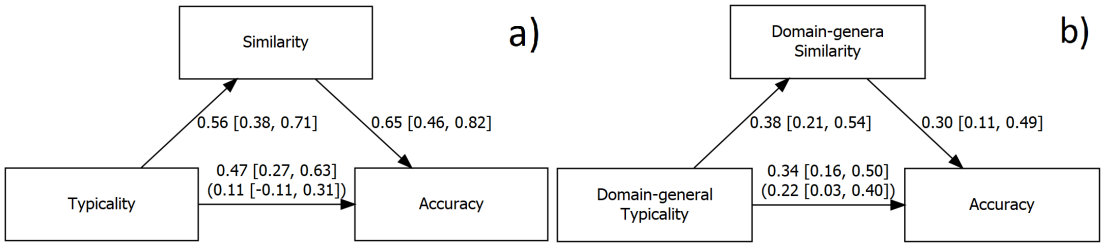

Figure B.8: Dyadic similarity mediated the Typicality-Accuracy link. Typical perceivers more accurately predicted novel targets' emotion transitions, and they did so by virtue of being more similar to their target.

Finally, we tested whether similarity and egocentrism were independent mechanisms that linked typicality to accuracy. In a multiple mediation model, we replicated the finding that similarity scores and egocentrism scores simultaneously mediated the association between typicality and accuracy. The total effect was $b = .42$ ($95\%CI = [.22, .59]$). We found positive indirect effects mediated through both egocentrism was ($b = .21, 95\%CI = [.03, .41]$) and similarity ($b = .28, 95\%CI = [.15, .42]$). After mediation, the effect of typicality was reduced to $b = -.07$ ($95\%CI = [-.31, .19]$). This pattern of results extended to domain general measures of typicality and similarity, such that egocentrism scores and domain general similarity scores simultaneously mediated the association between domain general typicality and accuracy. The total effect of domain general typicality was .33 ($95\%CI = [.15, .50]$). We found positive indirect effects through both egocentrism ($b = .16, 95\%CI = [.06, .27]$) and similarity ($b = .10, 95\%CI = [.02, .19]$). The effect of domain general typicality was reduced to $b = .08$ ($95\%CI = [-.10, .27]$; Figure S11).

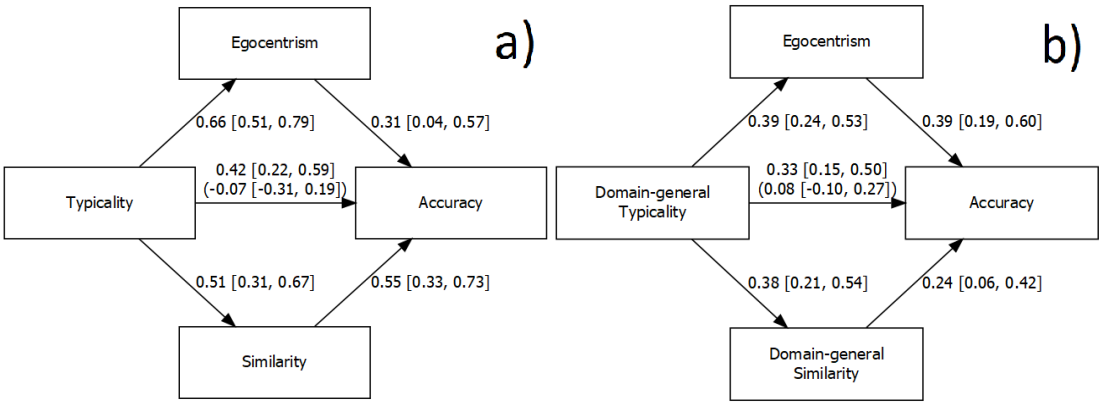

Figure B.9: Dyadic similarity and egocentric inference both independently mediated the Typicality-Accuracy link.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, "Science Experiment 02", was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.